

SKOLEEKSAMEN MET4



Høst 2023

Dato: 29. november 2023

Tidsrom: 09:00 - 12:00

Antall timer: 3

Foreleser/emneansvarlig kan kontaktes av eksamensvakt på telefon: 99385583

TILLATTE HJELPEMIDLER:

Kalkulator Ja ☒ Nei ☐

Ordbok: én tospråklig ordbok tillatt.

Alle trykte/egenskrevne hjelpemidler.

Antall sider, inkludert forside og vedlegg: 6

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

Oppgave 1

I en spørreundersøkelse¹ fra 2019 blant 500 tilfeldige nordmenn med høgskole- eller universitetsutdanning ble kjønn og om vedkommende hadde kort eller lang utdanning notert. Med kort utdanning mener vi universitets- og høgskoleutdanning på mindre enn eller lik 4 år. Med lang utdanning mener vi universitets- og høgskoleutdanning på mer enn 4 år. Resultatet er oppsummert i tabellen under:

	kort	lang
Menn	132	82
Kvinner	222	64

- (a) **Test om andelen med lang utdanning blant kvinner er forskjellig fra andelen med lang utdanning blant menn.**

Fordelingen av kjønn og lengde på utdanning blant personer med høgskole- eller universitetsutdanning i Norge var i 2009 som følger:

	kort	lang
Menn	0.315	0.150
Kvinner	0.439	0.096

- (b) **Test om fordelingen fra 2009 også kan stemme for 2019.**

Anta at vi har spurt n tilfeldige personer om de har høyere utdanning og Y av dem svarer ja. Da er $\hat{p} = Y/n$ et estimat på andelen av befolkningen med høyere utdanning. La X_i være en dummyvariabel som tar verdien 1 dersom individ nr. i i utvalget svarer at de har høyere utdanning og 0 ellers.

- (c) **Uttrykk estimatet \hat{p} som funksjon av variablene X_1, X_2, \dots, X_n og bruk dette resultatet til å argumentere for at \hat{p} er tilnærmet normalfordelt for stor n .**

Oppgave 2

En spørreundersøkelse² knyttet til bruk av Norefjell-Reinsjøfjellområdet (NR-området) ble sendt i en link per SMS til tre grupper: et utvalg av lokalbefolkningen i NR-området, et utvalg av fritidsboligeierne i NR-området, og et utvalg av innbyggere i Oslo og Viken. Respondentene tilhører kun en av disse gruppene. Videre ble respondentene bedt om å ta stilling til forskjellige scenarioer, blant annet et scenario med lite utbygging i NR-området.

Vedlegg 2 viser utskriften av en estimert logistisk regresjonsmodell som beskriver **sannsynligheten for at respondenter foretrekker scenariet med lite utbygging i NR-**

¹Disse dataene er simulert, men trukket fra en fordeling som samsvarer med de faktiske populasjonsandelene i 2019.

²<https://www.menon.no/wp-content/uploads/2022-41-Baerekraftig-utvikling-i-Norefjell-Reinsjofjell.pdf>

området som funksjon av gruppetilhørighet og andre forklaringsvariabler, se variabelbeskrivelse i Vedlegg 2.

- (a) Skriv opp den estimerte modellen.
- (b) Gi en fortolkning av den estimerte modellen.
- (c) Hva er sannsynligheten for at en 37 år gammel kvinne med doktorgradsutdanning og hytte i området foretrekker lite utbygging (ifølge modellen)?
- (d) Hvor gammel må kvinnen i oppgave c) være for at sannsynligheten for at hun foretrekker lite utbygging skal være 0.8 (ifølge modellen)?

Oppgave 3

Anta at den gjennomsnittlige daglige temperaturen i Bergen ved dag t kan skrives på følgende måte:

$$\text{temperatur}_t = T_t + S_t + R_t,$$

der T_t er en trendkomponent, S_t er en sesongkomponent og R_t er residualserien, altså det som ikke fanges opp av trend- og sesongkomponentene. I Vedlegg 3 ser vi tilsvarende estimerte trend- og sesongkomponenter for den gjennomsnittlige daglige temperaturen i Bergen. Vedlegg 3 viser også utskriften av en modell som har blitt tilpasset residuالتidsrekken R_t .

- (a) Forklar *kort* hva vi lærer av å se på de estimerte trend- og sesongkomponentene.
- (b) Hvilken tidsrekkemodell har vi tilpasset residualserien R_t ? Skriv opp modellen som har blitt tilpasset residualserien. Dersom dette er den sanne modellen for R_t , er R_t stasjonær?
- (c) Verdien av residualserien ved tid t er $R_t = 0.25$. Prediker R_{t+1} ved hjelp av modellen for residualserien. De tilsvarende prediksjonene for trend- og sesongkomponenten er $\hat{T}_{t+1} = 8.40$ og $\hat{S}_{t+1} = -4.80$. Hva er prediksjonen av temperatur ved tid $t + 1$?
- (d) Prediksjonene for trend- og sesongkomponenten ved tid $t + 365$ (ett år frem i tid) er $\hat{T}_{t+365} = 8.40$ og $\hat{S}_{t+365} = -4.80$. Hva er prediksjonen av R_{t+365} og temperaturen om ett år? Kommenter.

Vedlegg 1: Diverse R-utregninger

```
qnorm(0.975)
```

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

```
qnorm(0.95)
```

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

```
qnorm(0.90)
```

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

```
qchisq(.95, df = 2)
```

```
## [1] 5.991465
```

```
qchisq(.95, df = 3)
```

```
## [1] 7.814728
```

```
qchisq(.95, df = 4)
```

```
## [1] 9.487729
```

Vedlegg 2: Logistisk regresjon for spørreundersøkelse

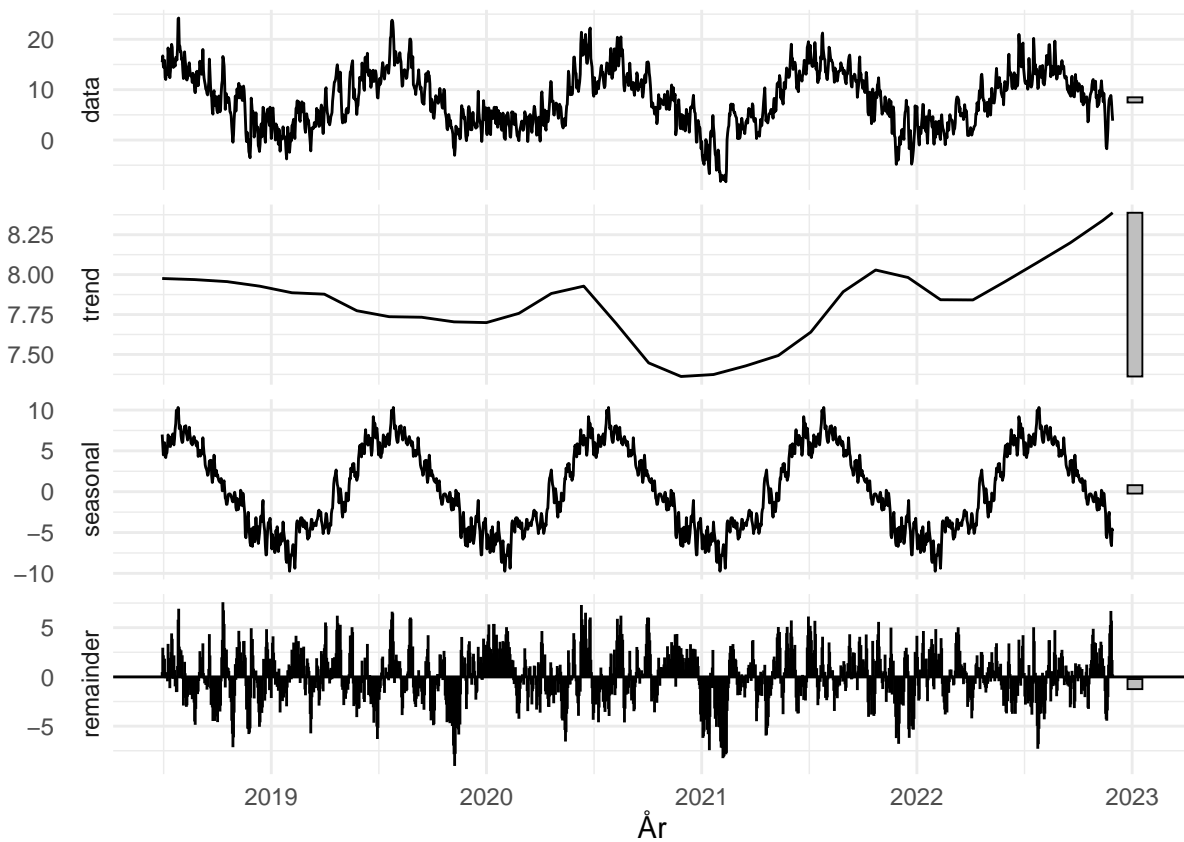
Tabell 1: Variabelbeskrivelser.

foretrekkerLU	Dummyvariabel som tar verdien 1 dersom respondenten foretrekker lite utbygging
fritidsbeboer	Dummyvariabel som tar verdien 1 dersom respondenten er fritidsbeboer i NR-området
lokalbeboer	Dummyvariabel som tar verdien 1 dersom respondenten er lokalbeboer i NR-området
alder	Alder på respondenten
utdanning	Utdanningsnivå til respondenten på en skala fra 1-6. 1 = Grunnskole, 2 = Videregående skole, 3 = Fagbrev, 4 = 3-4år universitet, 5 = 5årig universitet, 6 = doktorgrad.
kvinne	Dummyvariabel som tar verdien 1 dersom respondenten er kvinne

Estimert modell:

```
##
## =====
##                               Dependent variable:
##                               -----
##                               foretrekkerLU (logit)
## -----
## fritidsbeboer                0.078
##                               (0.130)
##
## lokalbeboer                 -0.346**
##                               (0.139)
##
## alder                       0.017***
##                               (0.004)
##
## utdanning                   0.089*
##                               (0.048)
##
## kvinne                      0.518***
##                               (0.109)
##
## constant                    -0.583**
##                               (0.290)
##
## -----
## Observations                 1,748
## Log Likelihood               -1,043.919
## Akaike Inf. Crit.           2,099.838
## =====
## Note:                       *p<0.1; **p<0.05; ***p<0.01
```

Vedlegg 3: Tidsrekkemodell for temperatur



```
## Series: R
## ARIMA(1,0,0) with zero mean
##
## Coefficients:
##      ar1
##      0.8147
## s.e.  0.0144
##
## sigma^2 = 2.357:  log likelihood = -2980.22
## AIC=5964.44   AICc=5964.45   BIC=5975.21
```