

HJEMMEEKSAMEN MET4



Vår 2020

Dato: 15. mai 2020

Tidsrom: 09:00 - 13:00

Antall timer: 4

BESVARELSEN SKAL LEVERES I WISEFLOW

På våre nettsider finner du informasjon om hvordan du leverer din besvarelse:
<https://www.nhh.no/for-studenter/eksamen/innlevering-individuelt-og-i-gruppe/>

Kandidatnummer blir oppgitt på StudentWeb i god tid før innlevering. Kandidatnummer skal være påført på alle sider øverst i høyre hjørne (ikke navn eller studentnummer). Ved gruppeinnlevering skal alle gruppemedlemmers kandidatnummer påføres.

Samarbeid mellom individer eller grupper om utarbeidelse er ikke tillatt, og utveksling av egenprodusert materiale til andre individer eller grupper skal ikke forekomme. En besvarelse skal bestå av individets, eller gruppens egne vurderinger og analyse. All kommunikasjon under hjemmeeksamen er å anse som fusk. Alle innleverte oppgaver blir behandlet i Urkund, NHHs datasystem for tekst- og plagiatkontroll

UTFYLLENDE BESTEMMELSER OM EKSAMEN

<https://www.nhh.no/globalassets/for-studenter/forskrifter/utfyllende-bestemmelser-til-forskrift-om-fulltidsstudiene-ved-nhh.pdf>

Antall sider, inkludert forside: 10

Antall vedlegg: 5 (Alle vedlegg følger etter oppgavene)

Oppgave 1

I forbindelse med kommunevalgkampen i 2019 var det fokus på utslipp av klimagasser. To av regjeringspartiene, Høyre og Venstre, ønsket å formidle budskapet om at Norges klimagassutslipp hadde gått ned under deres regjeringstid, og publiserte de to grafene som er vist i **Vedlegg 1a** og **1b**.

Begge disse figurene ble kritisert for å vere *misvisende*. På hvilke(n) måte(r) er de det? Beskriv kort med ord hvordan du tenker at figurene kunne vært mindre misvisende.

Oppgave 2

Lokal luftforurensing er et stort problem i mange byområder. Som ledd i en større kartlegging av luftkvaliteten i en stor europeisk by har myndighetene satt opp en sensor langs en travel innfartsåre som måler en rekke parametre hver time, deriblant konsentrasjonen av nitrogendioksid (NO_2). Dette er en gass som i store doser kan føre til svekket lungefunksjon og forverring av astma og bronkitt.

Vi skal i denne oppgaven undersøke data fra denne måleren, og ser på den gjennomsnittlige daglige konsentrasjonen av NO_2 , målt over en periode på 391 dager. I første omgang ønsker vi å se om det er forskjell i forventet NO_2 -konsentrasjon mellom helgedager (lørdag og søndag) og ukedager (mandag til fredag). I tabellen under finner vi en deskriptiv statistikk for målingene fordelt på de to kategoriene. Måleenheten er mikrogram per kubikkmeter ($\mu\text{g}/\text{m}^3$).

	Gj. snitt	St. avvik	Median	Min	Max	N
Ukedager	116.8	31.0	113.2	43.6	223.2	279
Helg	99.0	32.6	91.8	38.0	215.4	112

a) Test om variansen til NO_2 -konsentrasjonen er lik mellom ukedager og helgedager.

b) Test om forventet NO_2 -konsentrasjon er lik mellom ukedager og helgedager.

c) Hvilke forutsetninger gjør vi for å gjennomføre testene i spørsmål a) og b)?
Bruk informasjonen du har tilgjengelig til å vurdere om forutsetningene er oppfylt.

Dersom konsentrasjonen av NO_2 overstiger $100 \mu\text{g}/\text{m}^3$ er det ikke anbefalt at barn, eldre, eller personer med lungesykdommer oppholder seg utendørs i lengre perioder, og myndighetene må utstede et såkalt gult farevarsel. Vi ønsker videre å analysere om det er systematiske forskjeller også mellom ukedagene. Vi får oppgitt kontingenstabellen under, som viser antall dager med gult farevarsel, fordelt på de forskjellige ukedagene.

Ukedag	Antall dager med gult farevarsel	Antall dager uten gult farevarsel
Mandag	31	20
Tirsdag	34	17
Onsdag	38	15
Torsdag	36	16
Fredag	42	9

d) Test om kjennetegnene “gult farevarsel” og “ukedag” er uavhengige. Hva betyr resultatet i praksis? Du får oppgitt at testobservatoren i den aktuelle testen er gitt ved $\chi^2 = 6.14$ (Du trenger altså ikke skrive opp hele utregningen av testobservatoren, det holder at du viser hvordan det kan gjøres).

For å bedre forstå hvilke faktorer som forklarer variasjon i NO₂-konsentrasjon setter vi opp en regresjonsmodell basert på et datasett med følgende forklaringsvariabler:

Variabel	Forklaring
WeekdayMonday - WeekdaySunday	Dummyer som angir dag
Temperature	Daglig gjennomsnittlig lufttemperatur (°C)
Humidity	Daglig gjennomsnittlig relativ luftfuktighet (%-poeng)
Winter	Dummy som tar verdien 1 fra oktober t.o.m. mars, 0 ellers

I **Vedlegg 2** finner du en lineær regresjonsmodell med NO₂-konsentrasjonen som responsvariabel i kolonne (1).

e) For en gitt årstid, temperatur og luftfuktighet, hvilken ukedag har i følge den estimerte regresjonsmodellen høyest forventet NO₂-konsentrasjon?

f) Gi en *kortfattet* fortolkning av regresjonsmodell (1) i Vedlegg 2.

g) Bruk figurene i Vedlegg 3 til å diagnostisere regresjonsmodell (1) i Vedlegg 2. Skriv kortfattet.

Vi tilpasser en ny regresjonsmodell med de samme forklaringsvariablene, men denne gangen bruker vi logistisk regresjon, og som responsvariabel bruker vi dummyvariabelen *danger_warning*, som indikerer om gjennomsnittskonsentrasjonen av NO₂ den aktuelle dagen oversteg 100 µg/m³, og at myndighetene derfor måtte utstede gult farevarsel. Den estimerte modellen er gitt i kolonne (2) i **Vedlegg 2**.

h) På hvilken måte gir den logistiske regresjonen et annet bilde enn resultatet vi fikk i oppgave d)? Hvordan forklarer du det?

Myndighetene må bestemme seg for om de skal utstede farevarsel et døgn i forveien. I morgen er det lørdag 16. mai, og i den aktuelle byen er det meldt en gjennomsnittlig temperatur på 19 °C og en gjennomsnittlig relativ luftfuktighet på 47%.

i) Bruk den logistiske regresjonsmodellen til å predikere *sannsynligheten* for at gjennomsnittlig NO₂-konsentrasjon overstiger 100 µg/m³. Gi en kort vurdering

om myndighetene bør utstede gult farevarsel. (Husk at luftfuktigheten er gitt på skala 0–100, og ikke 0–1)

j) Svar spørsmål i) ved å benytte den lineære regresjonsmodellen i Vedlegg 2, kolonne (1) i stedet. Se bort fra usikkerhet knyttet til estimering av regresjonskoeffisientene når du svarer på dette spørsmålet.

Oppgave 3

Vi tenker oss at konsentrasjonen av NO_2 ved tidspunkt t kan skrives på følgende måte:

$$\text{NO}_{2t} = T_t + S_t + R_t,$$

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 4** ser vi et plott av NO_{2t} , samt estimerte trend- og sesongkomponenter, og et autokorrelasjonsplott for residualtidsrekken.

a) Forklar *kort* hva vi lærer av å se på de estimerte trend- og sesongkomponentene.

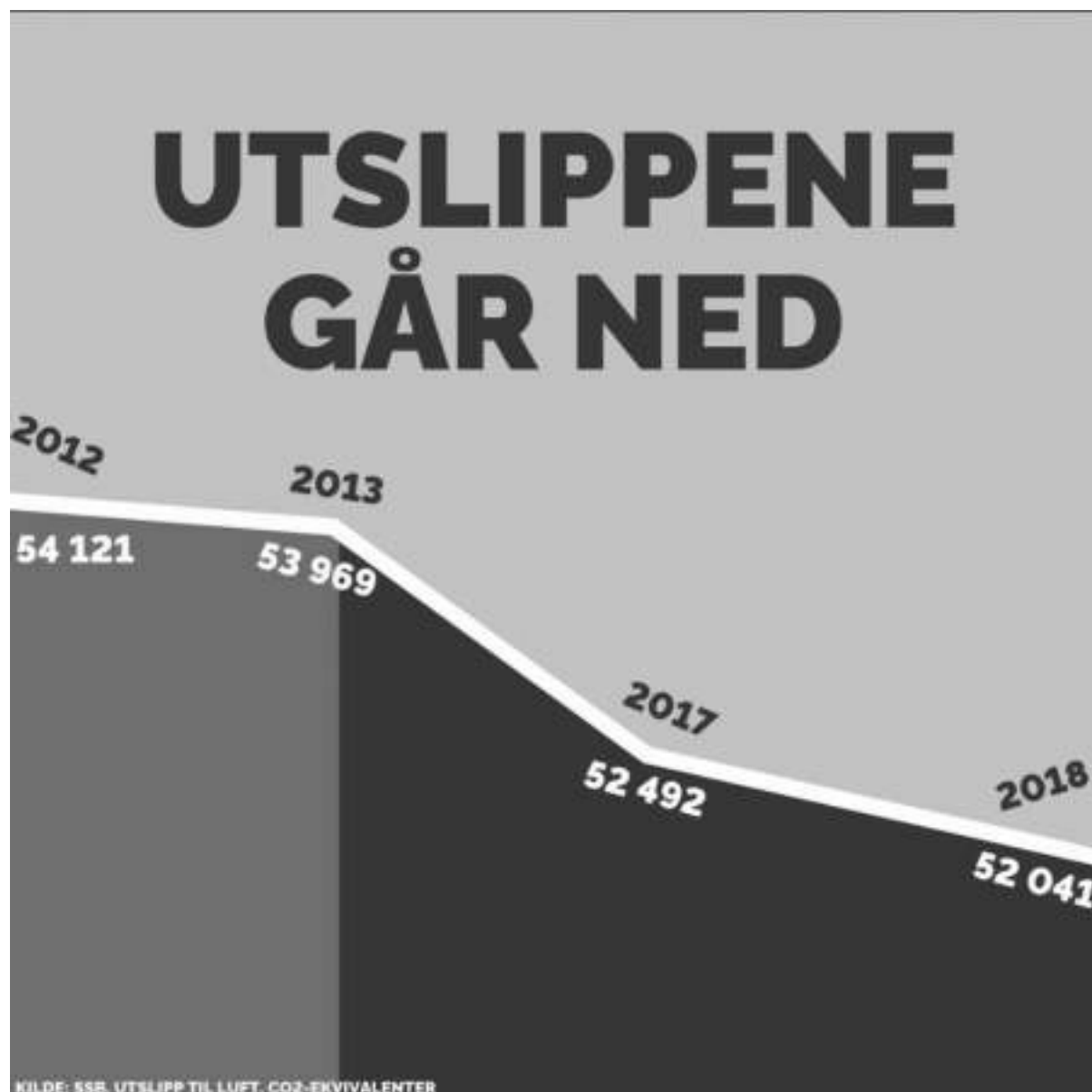
I **Vedlegg 5** har vi tilpasset to ulike tidrekkemodeller til residualtidsrekken R_t .

b) Hvilke to tidsrekkemodeller har vi tilpasset? Hvilken av de to modellene passer best til datasettet? Begrunn svaret, både ved hjelp av utskriftene i **Vedlegg 5** og en av figurene i **Vedlegg 4**.

Anta at den beste modellen fra spørsmål b) representerer den *sanne* modellen for residualtidsrekken R_t .

c) Er R_t stasjonær? Begrunn svaret.

Vedlegg 1a: Graf publisert av Høyre



Bakgrunn: Denne grafen ble publisert av Høyre på Facebook 1. november 2019, og viser norske CO₂-utslipp (i 1000 tonn CO₂-ekvivalenter) som funksjon av tid. Høyre overtok regjeringsmakten sammen med Fremskrittspartiet etter stortingsvalget i 2013.

Vedlegg 1b: Graf publisert av Venstre

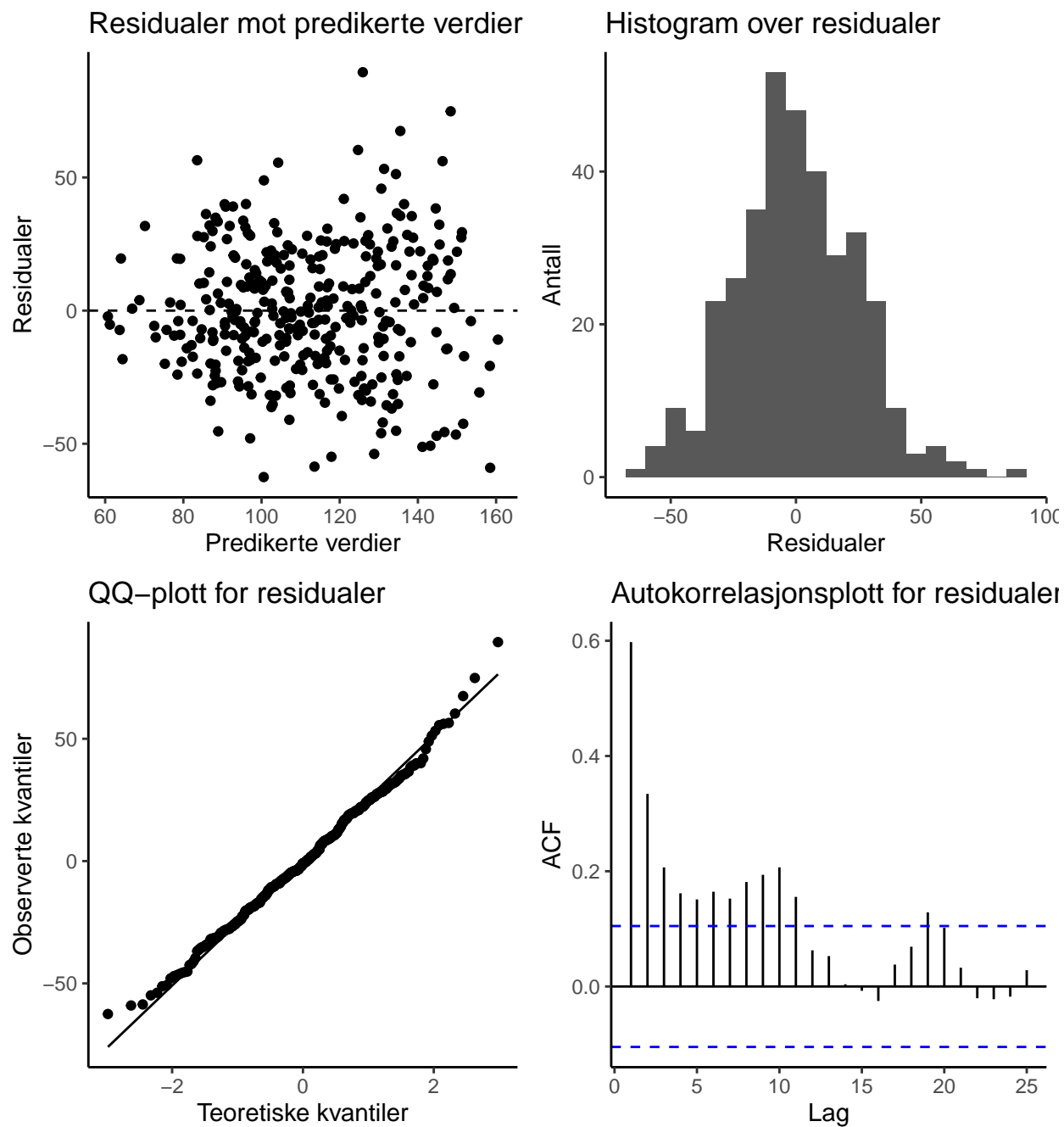


Bakgrunn: Denne grafen ble publisert av klima- og miljøminister Ola Elvestuen fra Venstre på Twitter 1. november 2019 (men senere tatt bort), og viser norske CO₂-utslipp (i 1000 tonn CO₂-ekvivalenter) som funksjon av tid. Venstre gikk inn i regjering sammen med Høyre og Fremskrittspartiet i januar 2018.

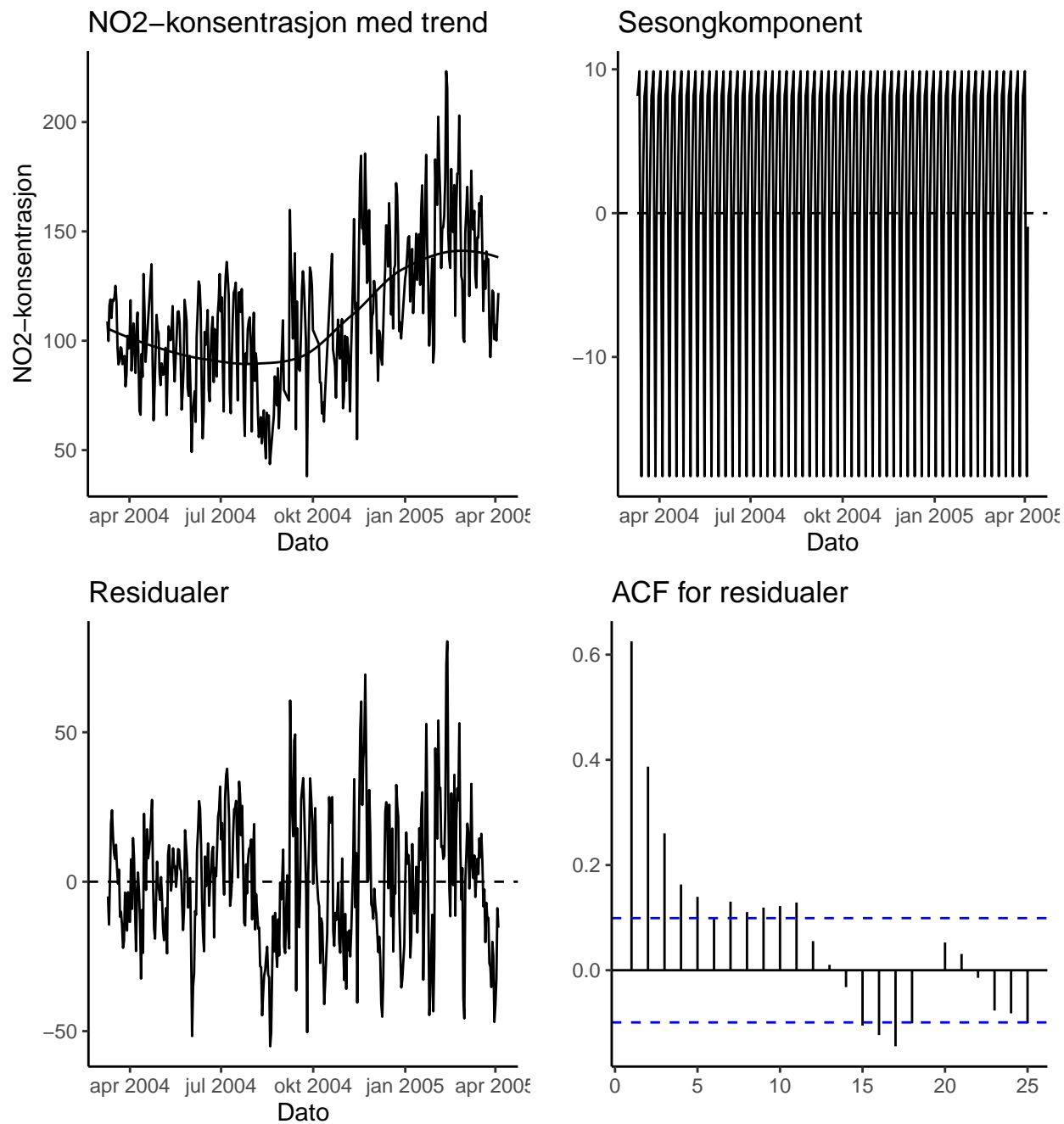
Vedlegg 2: Regresjonsutskrifter

Dependent variable:		
	no2 OLS (1)	danger_warning logistic (2)
WeekdayMonday	-16.562*** (4.958)	-1.587*** (0.521)
WeekdaySaturday	-19.239*** (4.889)	-2.292*** (0.519)
WeekdaySunday	-35.619*** (5.050)	-2.730*** (0.545)
WeekdayThursday	-4.949 (4.930)	-0.915* (0.521)
WeekdayTuesday	-11.389** (4.946)	-1.147** (0.523)
WeekdayWednesday	-5.857 (4.909)	-0.825 (0.525)
Temperature	-1.787*** (0.252)	-0.086*** (0.025)
Humidity	-0.517*** (0.121)	-0.044*** (0.013)
Winter	17.688*** (4.187)	1.289*** (0.393)
Constant	172.893*** (9.485)	5.052*** (1.043)
Observations	349	349
R2	0.431	
Adjusted R2	0.416	
Log Likelihood		-185.570
Akaike Inf. Crit.		391.141
Residual Std. Error	24.640 (df = 339)	
F Statistic	28.489*** (df = 9; 339)	
Note: *p<0.1; **p<0.05; ***p<0.01		

Vedlegg 3: Diagnoseplott til regresjon (1) i Vedlegg 2



Vedlegg 4: Dekomponering av tidsrekke



Forklaring:

- Oppe til venstre er NO₂-konsentrasjonen plottet gjennom observasjonsperioden sammen med en estimert trendkomponent.
- Oppe til høyre har vi plottet sesongkomponenten, som har en periode på 7 dager.
- Nede til venstre har vi plottet NO₂-konsentrasjonen etter at vi har trukket ut trend- og sesongkomponentene fra tidsrekken ("residualene").
- Nede til høyre ser vi den estimerte autokorrelasjonsfunksjonen til residualtidsrekken.

Vedlegg 5: To estimerte tidsrekkemodeller for Rt

```
Call:
arima(x = airquality_ts$residuals, order = c(1, 0, 0))

Coefficients:
      ar1  intercept
      0.6245    -0.0821
s.e.    0.0393     2.3389

sigma^2 estimated as 304.1:  log likelihood = -1672.83,  aic = 3351.65

Training set error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE
Training set 0.009145292 17.43983 13.27842 -133.9213 426.148 0.9166486
              ACF1
Training set 0.00558858
```

```
Call:
arima(x = airquality_ts$residuals, order = c(0, 0, 1))

Coefficients:
      ma1  intercept
      0.5536    -0.0151
s.e.    0.0391     1.4523

sigma^2 estimated as 342.3:  log likelihood = -1695.86,  aic = 3397.73

Training set error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE
Training set -0.004784311 18.50129 14.31871 14.36084 241.087 0.9884633
              ACF1
Training set 0.1395284
```