

HJEMMEEKSAMEN MET4



Vår 2022

Dato: 6. mai 2021

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 og vedlegg: 10

Antall vedlegg: 4 (Alle vedlegg følger etter oppgavene)

Introduksjon

PISA (Programme for International Student Assessment) er et prosjekt i regi av OECD som måler evner hos skolebarn på tvers av land ved hjelp av standardiserte tester. Etter at den første PISA-testen ble gjennomført i år 2000 har resultatene hatt stor innvirkning på utdanningspolitikken i mange land. I en studie fra 2019¹ kan vi lese at Finlands overraskende gode resultater i den første testen gjorde at de ble sett på som et mønsterland og eksempel til etterfølgelse, mens Tyskland, for eksempel, innførte drastiske tiltak for å forbedre utdanningssystemet sitt etter dårlige resultater.

Den samme studien stiller derimot et viktig spørsmål: Måler slike standardiserte tester bare forskjell i evnenivå mellom land, eller kan det oppstå forskjeller som følge av andre årsaker, som for eksempel at elevene i ulike land har ulike insentiver for å gjøre det bra på testen?

For å undersøke dette gjennomførte forskergruppen følgende eksperiment: En gruppe elever fra USA (et land som generelt gjør det dårlig på PISA-tester) og en gruppe elever fra Kina (som gjør det svært bra på PISA-testene) gjennomførte den samme standardiserte prøven i matematikk. Prøven hadde 25 spørsmål. Innen hvert land ble så elevgruppen tilfeldig delt i to omtrent like store grupper. Den ene gruppen (kontrollgruppen) gjennomførte testen på vanlig måte uten noen spesielle insentiver til å gjøre det bra. Den andre gruppen (behandlingsgruppen) fikk utlevert en konvolutt med \$25 (og en tilsvarende pengesum for de kinesiske elevene i lokal valuta) rett før prøven ble gjennomført, og fikk vite at de kunne beholde pengene, men at de ville miste \$1 for hvert spørsmål de svarte feil eller ikke svarte på.

Kontrollgruppen og behandlingsgruppen var ikke klar over hverandre, og behandlingsgruppen visste ikke om belønningssystemet før de møtte opp til prøven, så de hadde ikke insentiver til å forberede seg på annen måte enn kontrollgruppen.

Vi har dermed fire elevgrupper, en kontrollgruppe og en behandlingsgruppe i henholdsvis USA og Kina. Vi samler inn poengsum (0–25 poeng) for hver elev, og får følgende deskriptive statistikk for hver av de fire gruppene:

Land	Gruppe	Min	Gj.snitt	Median	Max	St.avvik	N
Kina	Ikke mottatt insentiv (kontrollgruppe)	9	20.50	21.0	25	2.95	333
Kina	Mottatt insentiv (behandlingsgruppe)	7	20.23	21.0	25	3.52	323
USA	Ikke mottatt insentiv (kontrollgruppe)	1	10.22	10.0	23	5.64	227
USA	Mottatt insentiv (behandlingsgruppe)	0	12.15	11.5	25	6.00	220

Oppgave 1

- Test om variansen i kontrollgruppen er lik variansen i behandlingsgruppen. Gjennomfør testen både for USA og for Kina.
- Test om forventet poengscore er forskjellig mellom kontroll- og behandlingsgruppen. Gjennomfør testen både for USA og for Kina. Kommenter resultatet.

¹Uri Gneezy, John A. List, Jeffrey A. Livingston, Xingdong Qin, Sally Sadooffog Yang Xu: *Measuring Success in Education: The Role of Effort on the Test Itself*, American Economic Review: Insights, 2019

Elevene fra Kina er hentet fra fire ulike skoler i landets største by Shanghai. Vi ønsker å finne ut om det er statistisk signifikante forskjeller i evnenivå på tvers av skolene. Vi fordeler alle elevene i kontrollgruppen etter hvilken skole de går på (`school`) og poengintervall (`score_group`), og får følgende kontingenstabell:

school		score_group			
		[5,10]	(10,15]	(15,20]	(20,25]
shanghai	school 1	0	7	18	5
shanghai	school 2	0	6	49	50
shanghai	school 3	1	2	31	99
shanghai	school 4	1	2	24	38

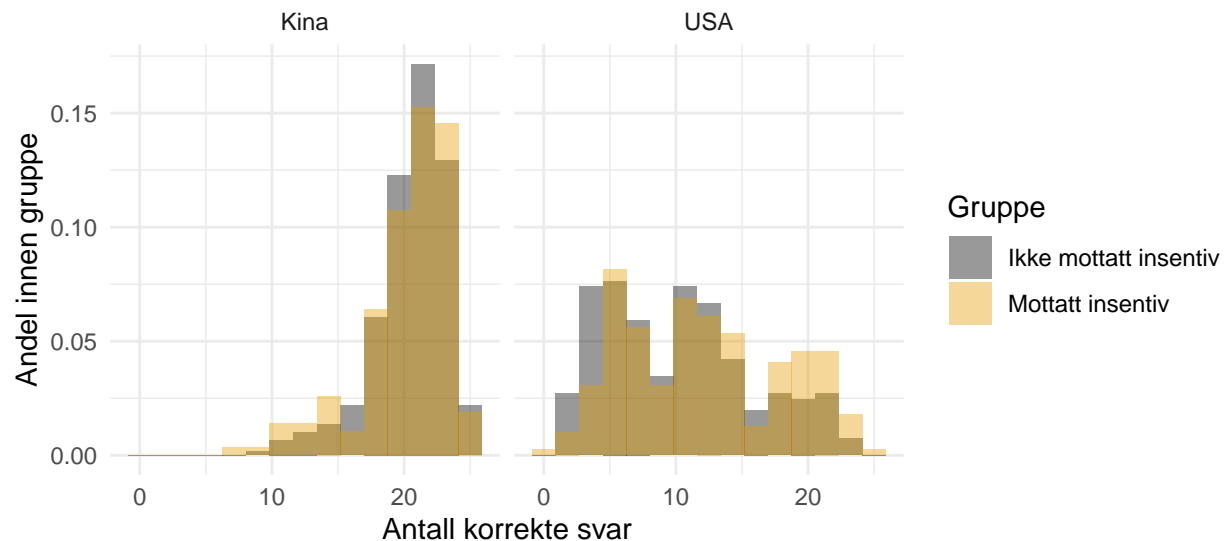
Vi gjennomfører en χ^2 -test på denne tabellen og får ut følgende utskrift:

Pearson's Chi-squared test

```
data:  tab_cn
X-squared = 56.2, df = 9, p-value = 7.188e-09
```

(c) Hva er det vi har testet her? Sett opp null- og alternativhypotesene, skisser utregning av testobservator og konkluder.

I figuren under ser du fire histogrammer for testresultatet, et for hver gruppe av elever. Histogrammene for hvert land er lagt over hverandre, med Kina til venstre og USA til høyre. De grå stolpene representerer kontrollgruppene, og de oransje stolpene representerer behandlingsgruppene.



(d) Bruk tabellen med deskriptiv statistikk og histogrammene i figuren over til å beskrive fordelingen av testresultat i USA og Kina, samt effekten av insentiver. Vurder spesielt om testresultatet er normalfordelt i de ulike gruppene og om konklusjonen din på dette spørsmålet har konsekvenser for konklusjonen i spørsmål (b)

Oppgave 2

Vi ønsker å justere resultatene våre for eventuelle demografiske forskjeller mellom den amerikanske og kinesiske elevgruppen, og setter opp følgende lineære regresjonsmodell:

$$\text{score} = \beta_0 + \beta_1 \cdot \text{treatment} + \beta_2 \cdot \text{female} + \beta_3 \cdot \text{age} + \epsilon,$$

der **score** er resultatet på testen, **treatment** er en dummyvariabel som tar verdien 1 dersom eleven er i behandlingsgruppen og 0 ellers, **female** er en dummyvariabel dersom eleven er en jente og 0 ellers, **age** er alderen til eleven, og ϵ er feilleddet i regresjonen.

I Vedlegg 1 har vi estimert denne regresjonsmodellen to ganger, en gang for de amerikanske elevene (kolonne 1) og en gang for de kinesiske elevene (kolonne 2).

- (a) Gi en kort fortolkning av utskriften i Vedlegg 1, og bruk regresjonsutskriften til å kommentere hypotesen om at elevgrupper responderer ulikt på insentiver.
- (b) Kommenter diagnoseplottene for regresjonsmodell 2 (Kina) som du finner i Vedlegg 2.
- (c) Prediker testscore for en 16 år gammel amerikansk gutt i behandlingsgruppen. Hva er fortolkningen av et prediksjonsintervall for denne prediksjonen? (Du skal ikke regne ut prediksjonsintervallet).
- (d) Lag et 95% konfidensintervall for koeffisienten til **treatment** i regresjonsmodellen som er estimert ved hjelp av amerikanske data.
- (e) Vi ser at effekten av insentiver er statistisk signifikant forskjellig fra null i USA, men *ikke* i Kina. Forklar at dette ikke nødvendigvis betyr at det er statistisk signifikant forskjell i effekt mellom de amerikanske og kinesiske elevene i studien. Sett opp en ny regresjonsmodell der vi kan direkte måle forskjellen i reaksjon på insentiver mellom amerikanske og kinesiske elever.

Oppgave 3

På en muntlig eksamen ved en annen institusjon blir en student presentert med et datasett bestående av det amerikanske private forbruket, målt i dollar hvert kvartal fra 1970 til 2016. Tidsrekken er plottet i Vedlegg 3. Studenten kommer så med følgende utsagn:

Studentens utsagn
Forbruket i USA er helt klart en ikke-stasjonær tidsrekke ettersom det er en tydelig stigende trend. Siden stasjonaritet er en forutsetning for Å¥ kunne modellere en tidsrekke kan vi ikke analysere disse dataene ved hjelp av tidsrekkeметoder.

(a) Hvordan ville du vurdert dette utsagnet dersom du var sensor? Begrunn svaret ditt.

Den prosentvise *forandringen* i det amerikanske private forbruket $\{Y_t\}$ er plottet i Vedlegg 4 sammen med den tilhørende empiriske autokorrelasjonen.

(b) Hva forteller denne autokorrelasjonsfunksjonen oss om tidsavhengigheten i *denne* tidsrekken?

Ved hjelp av en statistisk programvarepakke har vi funnet ut at følgende utskrift beskriver en modell som passer godt til $\{Y_t\}$:

```
Series: uschange[, "Consumption"]
ARIMA(1,0,3) with non-zero mean
```

Coefficients:

	ar1	ma1	ma2	ma3	mean
	0.5885	-0.3528	0.0846	0.1739	0.7454
s.e.	0.1541	0.1658	0.0818	0.0843	0.0930

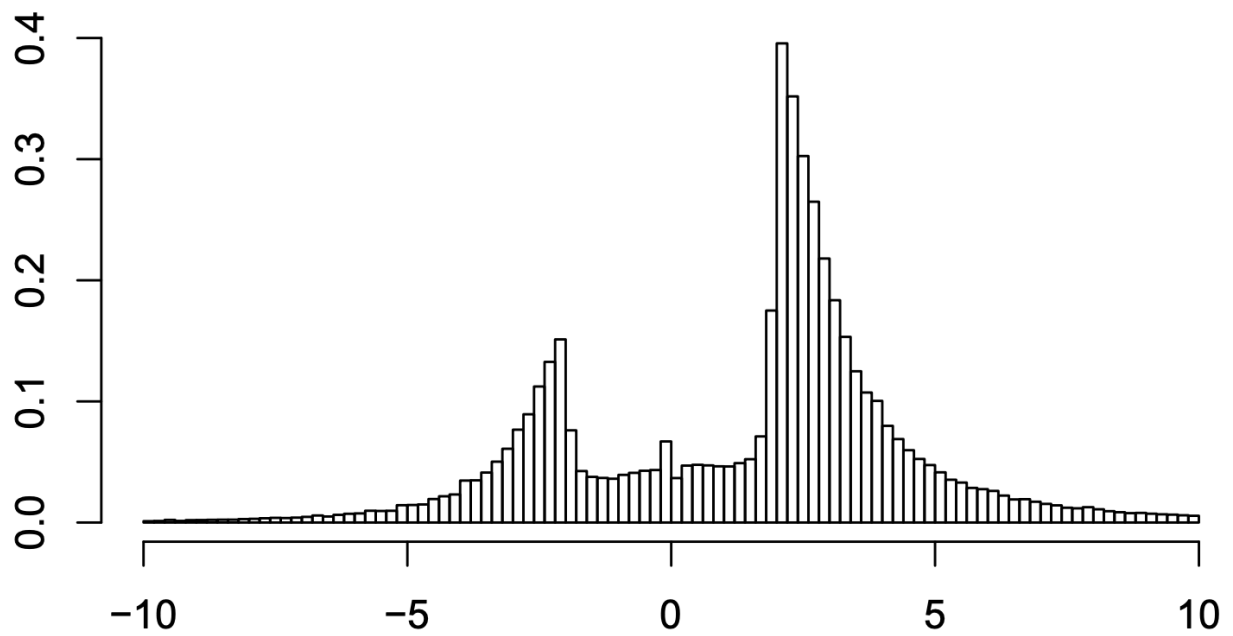
```
sigma^2 estimated as 0.3499: log likelihood=-164.81
AIC=341.61 AICc=342.08 BIC=361
```

(c) Hvilken modell er dette? Skriv den opp.

Oppgave 4

I en nylig publisert artikkel² har en nederlandsk forskergruppe samlet inn over *en million* testobservatorer fra *t*-tester som er gjort i medisinske forsøk fra 1976 til 2019 fra den anerkjente forskningsdatabasen Medline. Histogrammet for alle disse testobservatorene ser slik ut:

²van Zwet, EW, Cator, EA: *The significance filter, the winner's curse and the need to shrink*. Statistica Neerlandica, 2021.

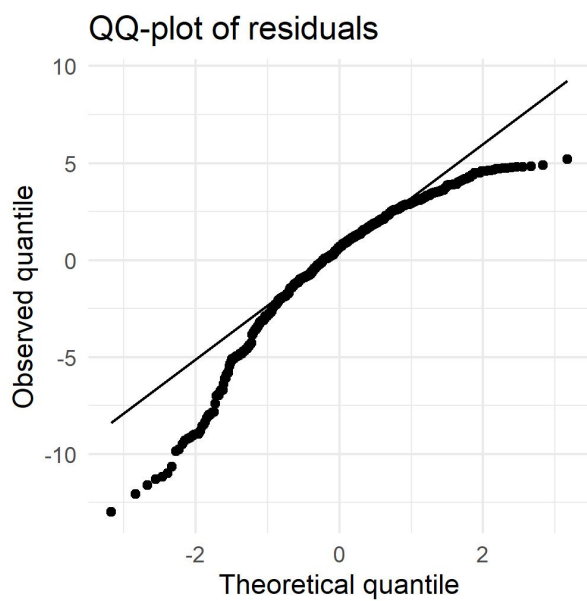
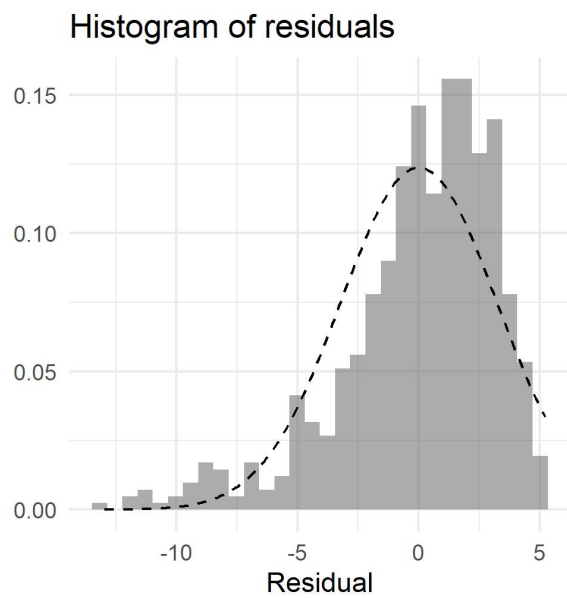
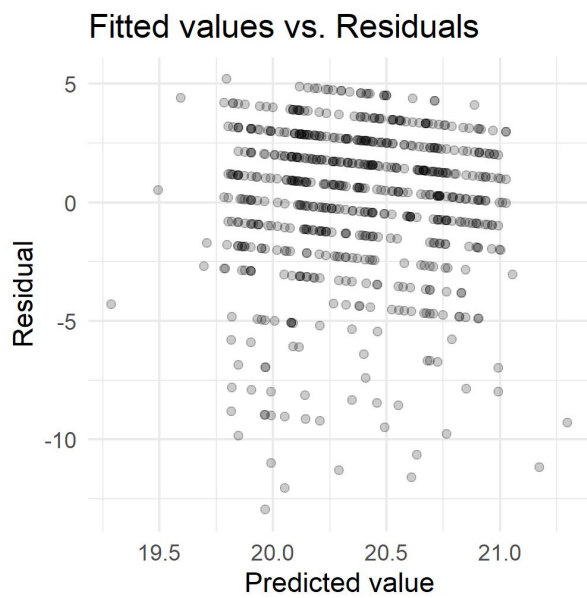


- (a) Histogrammet over har en spesiell form. Hva tror du er grunnen til at testobservatorene fordeler seg slik?

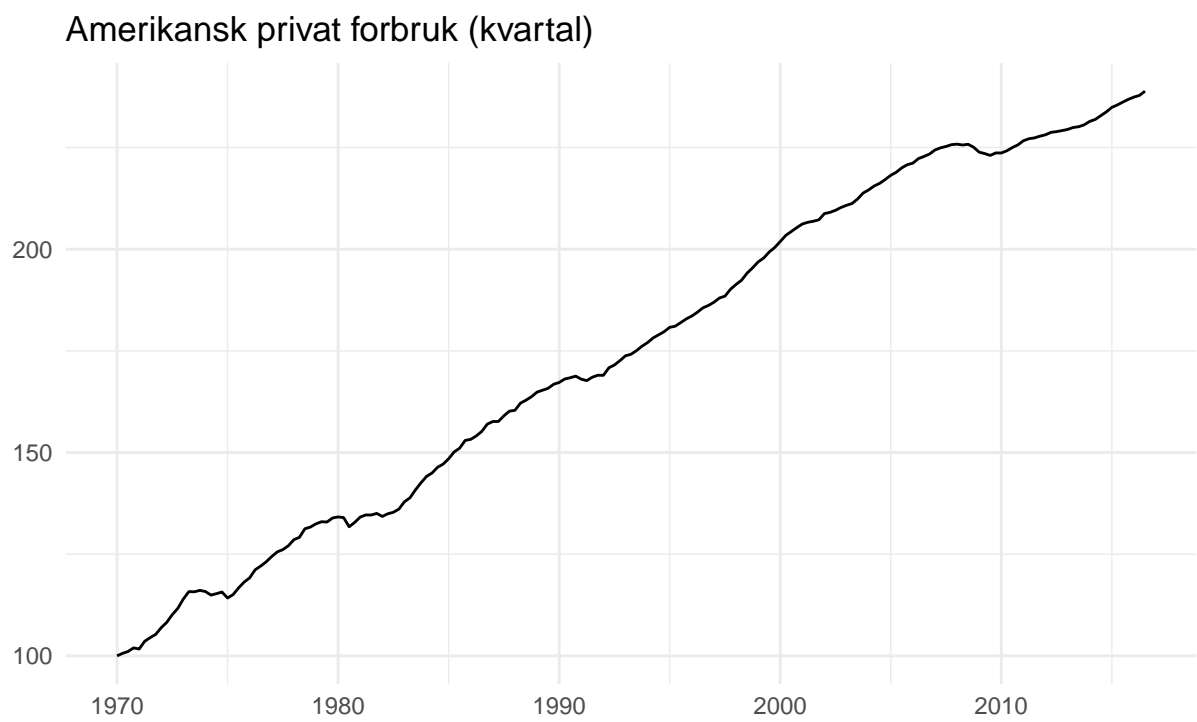
Vedlegg 1: Regresjonsmodeller

Dependent variable:		
	score	
	USA (1)	China (2)
treatment	1.897*** (0.552)	-0.300 (0.253)
age	-0.107 (0.390)	-0.350 (0.318)
female	-0.948* (0.552)	-0.586** (0.254)
Constant	12.421* (6.344)	26.507*** (5.169)
Observations	447	656
R2	0.033	0.011
Adjusted R2	0.027	0.007
Residual Std. Error	5.814 (df = 443)	3.233 (df = 652)
F Statistic	5.076*** (df = 3; 443)	2.466* (df = 3; 652)
Note: *p<0.1; **p<0.05; ***p<0.01		

Vedlegg 2: Residualplott

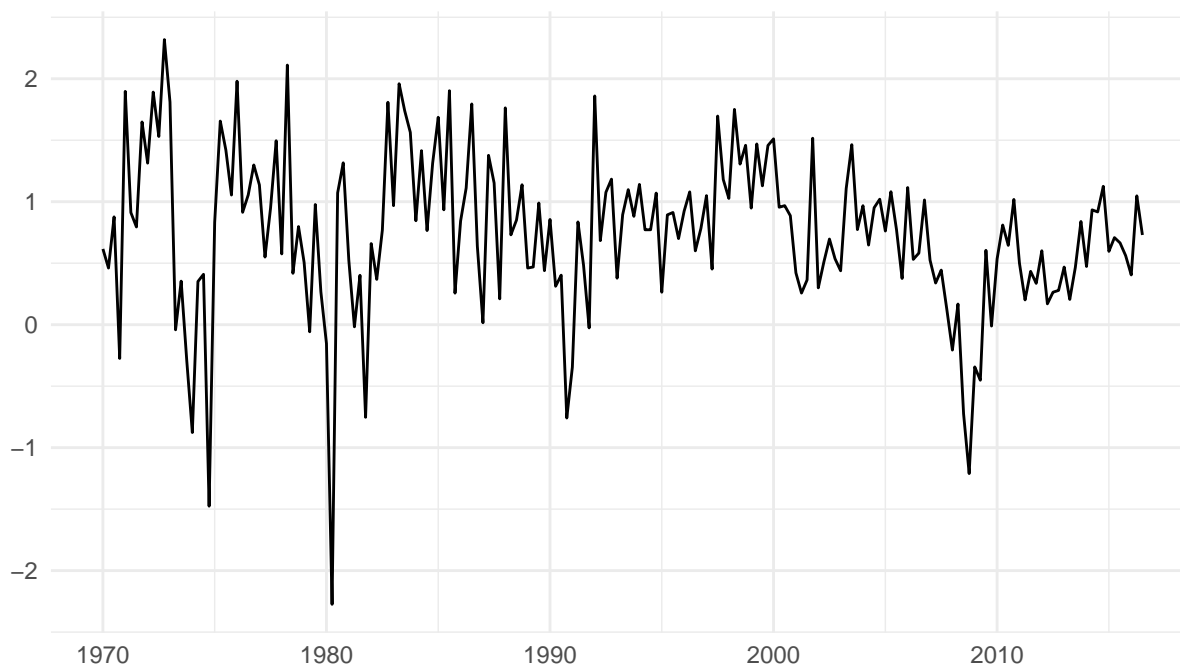


Vedlegg 3: Tidsrekke



Vedlegg 4: Tidsrekke, prosentvis avvik

Prosentvis forandring i amerikansk privat forbruk (kvartal)



Autokorrelasjonsfunksjonen

