

NORGES HANDELSHØYSKOLE

Eksamen vår 2013

Kurskode: INT010 **Tittel:** Anvendt metode

Dato: 7.6.2013 **Kl. 09.00-12.00**

Faglærer går ikke rundt i eksamenslokalene, men kan kontaktes av eksamensvakten på
tlf. 59612/99 69 93 05

Hjelpemidler til eksamen:

Hjelpemidler tillatt: JA, alle trykte/egenskrevne

Kalkulator: JA [I tråd med retningslinjer for bruk av kalkulator, jf. utfyllende bestemmelser til Forskrift om eksamen ved Norges Handelshøyskole (fulltidsstudiene).]

Alle delspørsmål i oppgavesettet teller likt.

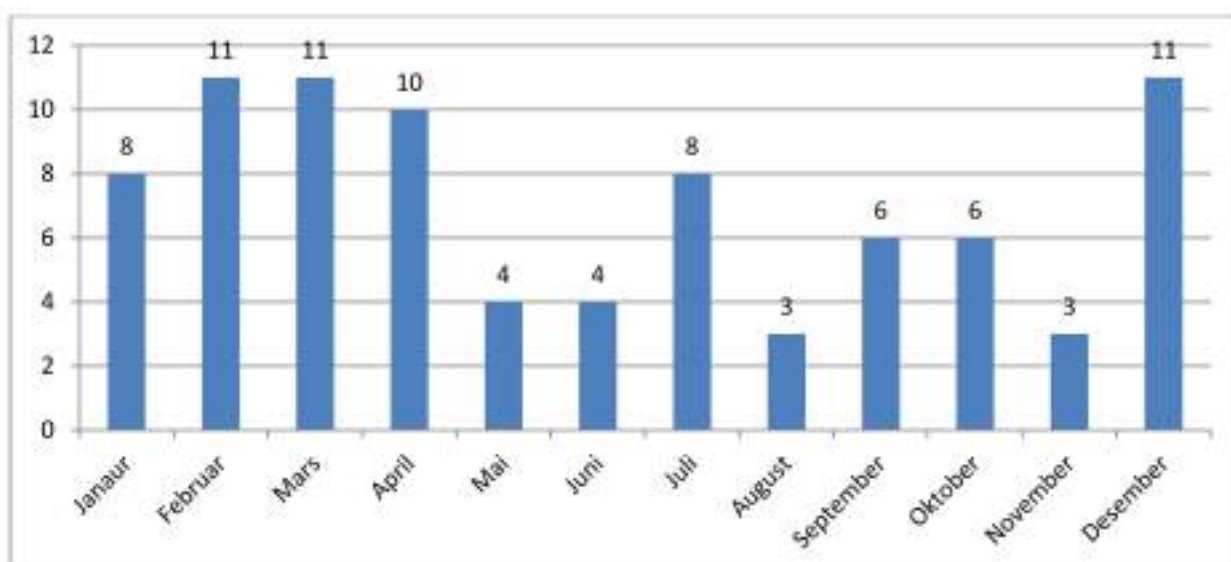
Oppgave 1

I en artikkel søndag 10. februar 2013 fokuserer BT på at barn som er født tidlig på året har større sannsynlighet for å ende opp som toppidrettsutøvere enn barn som er født sent på året.

I artikkelteksten kommer det fram at fenomenet er godt kjent og forsket på. BT gir tall fra flere idretter. Det virker som om det er vanlig å oppgi forholdet mellom prosent utøvere født i første og annet halvår. Eksempelvis står det at i eliteserien i fotballen «er fordelingen 60-40 i favør av dem som er født i årets seks første måneder».

Den mest sannsynlige forklaringen er at idretten definerer unge talenter ut fra hvem som er gode «der og da» og at dette ofte er de som er født tidlig på året. De er i gjennomsnitt sterkest og raskest fordi de er eldst innenfor sitt årskull.

Hovedillustrasjonen til BT er en opptelling av utøvere som har tatt individuelt gull i OL. Dataene er gjengitt i figuren nedenfor, og overskriften er at et stort flertall av norske OL-vinnere er født i første halvår.



- Test om forholdet mellom OL-vinnere født i første og annet halvår er signifikant forskjellig fra 50-50.
- Anta at du har planlagt å teste om andelen av de 85 OL-vinnerne som er født i desember er signifikant forskjellig fra $1/12$. Hvor mange OL-vinnere må da være født i denne måneden for at du skal påstå at andelen er signifikant forskjellig fra $1/12$ på 5 % signifikansnivå?
- Hvordan kan man argumentere for å bruke ensidig test i spørsmål b)? Bør alternativhypotesen ved ensidig test i denne konteksten være $p > 1/12$ eller $p < 1/12$?

Hvis man bare ser på en enkeltmåned eller aggregerer opp til halvår utnytter man ikke all informasjonen i datasettet. Dessuten er fødsler i normalbefolkningen heller ikke jevnt fordelt over året.

- d) Hvordan vil du teste om OL-vinnere har en annen fordeling av fødselsmåned enn resten av befolkningen ved å bruke dataene ovenfor i kombinasjon med demografiske data for norske fødsler? Sett opp formelen for testobservatoren og forklar kort hvordan du vil regne ut tallene som inngår i formelen.

I en oppfølgingsartikkel ser BT på Brannspillere. Det fremkommer at i U-stallen til Brann er «16 av 25 født i årets seks første måneder» og at fordelingen i A-stallen er «enda mer tydelig». Der er «15 er født i første halvår, 7 i andre halvår».

- e) Test om andelen av spillere som er født i første halvår er signifikant forskjellig i U-stallen og A-stallen.

Oppgave 2

Kunnskapsdepartementet jobber for tiden med en omlegging av pressestøtten. Et offentlig utvalg foreslo høsten 2009 at det nåværende momsfrirket for papiraviser delvis skal erstattes med et skattefradrag basert på avisenes redaksjonelle kostnader. Et konsultentselskap har fått i oppdrag å utrede hvordan dette vil slå ut og trenger en modell for å predikere størrelsen på avisenes redaksjonelle kostnader. De har hentet data fra nettsidene til Mediebedriftenes landsforbund og Norsk journalistlag.

De har laget et datasett med følgende variabler:

redkost	= redaksjonelle kostnader i 2010 i millioner kroner (beregnet som totale lønnskostnader til medlemmer av Norges journalistlag i hver avis multiplisert med en faktor på 1,4).
lnredkost	= $\ln(\text{redkost})$
opplag	= avisenes gjennomsnittlige opplag per utgave på hverdager i 2010 oppgitt i 1000 aviser.
lnopplag	= $\ln(\text{opplag})$
utgaver	= antall hverdager som avisen utkommer på
utg_oppl	= antall utgaver multiplisert med gjennomsnittlig opplag per utgave
no2	= dummyvariabel med verdi 1 dersom avisen er en «nummer 2-avis» i det lokale avismarkedet. (Eksempelvis er BA en nummer 2-avis i Bergen.)
oslo	= dummyvariabel med verdi 1 dersom avisen utgis fra Oslo

Deskriptiv statistikk for 128 aviser med komplett informasjon om opplag og kostnader er gitt under.

Summary statistics (128 valid observations)

	Mean	Std.Dev.	Median	Minimum	Maximum
redkost	14.875	27.549	5.565	0.623	200.492
lnredkost	1.928	1.151	1.716	-0.474	5.301
opplag	17.589	38.456	7.777	1.271	344.843
lnopplag	2.228	0.948	2.051	0.240	5.843
utgaver	4.344	1.736	5.000	1.000	6.000
utg_oppl	96.168	233.505	35.643	1.271	2069.060
no2	0.078	0.269	0.000	0.000	1.000
oslo	0.086	0.281	0.000	0.000	1.000

Correlation coefficients

redkost	lnredkost	opplag	lnopplag	utgaver	utg_oppl	no2	oslo	
1.000	0.759	0.943	0.815	0.400	0.947	-0.047	0.464	redkost
	1.000	0.628	0.946	0.738	0.643	-0.012	0.363	lnredkost
		1.000	0.726	0.298	0.998	-0.046	0.496	opplag
			1.000	0.643	0.732	-0.004	0.434	lnopplag
				1.000	0.337	-0.041	0.052	utgaver
					1.000	-0.046	0.479	utg_oppl
						1.000	0.015	no2
							1.000	oslo

(Korrelasjonen mellom to variabler finnes i krysset mellom variablenes rad og kolonne. Eksempelvis er korrelasjonen mellom *utgaver* og *opplag* 0,298)

- Hvor mange av avisene i utvalget blir utgitt fra Oslo?
- Kommenter kort sammenhengen mellom opplagsstørrelse og det å være nummer to-avis.

Konsulentselskapet merker seg den høye korrelasjonen mellom redaksjonelle kostnader og opplaget og begynner med en meget enkel modell:

$$\text{Redkost} = \alpha + \beta \cdot \text{opplag} + \varepsilon$$

- Bruk opplysningene i den deskriptive statistikken til å vise at regresjonskoeffisienten β blir 0,676 og α blir 2,985.
- Hva er forventet redaksjonell kostnad i en avis med gjennomsnittlig opplagsstørrelse?
- Standardavviket til feilleddet i modellen er 9,184. Regn ut et 95 prosent konfidensintervall for forventet redaksjonell kostnad i en avis med gjennomsnittlig opplagsstørrelse.

Utskriften under viser resultatet av en litt mer avansert regresjonsmodell.

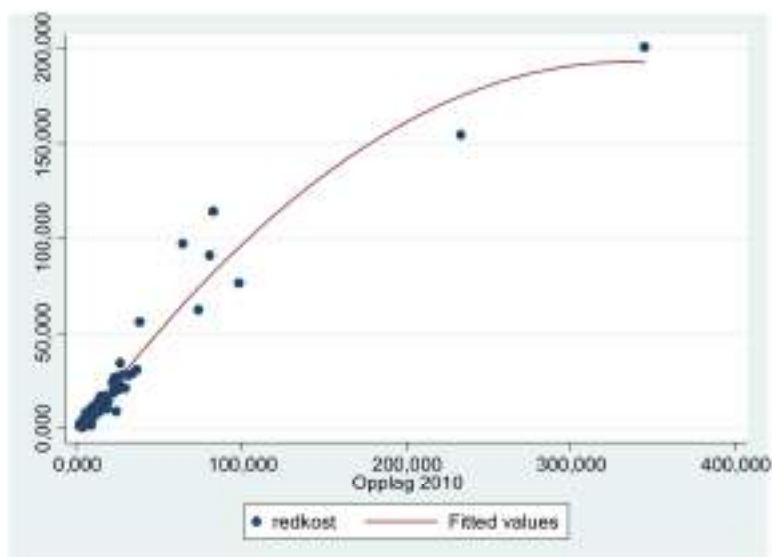
Model 2: OLS, using observations 1-128

Dependent variable: lnredkost

	coefficient	std. error	t-ratio	p-value	
const	-0.937198	0.0952525	-9.839	3.59e-017	***
lnopplag	1.01700	0.0592228	17.17	2.39e-034	***
utgaver	0.142511	0.0233782	6.096	1.32e-08	***
utg_oppl	-0.000237205	0.000191391	-1.239	0.2176	
no2	-0.00873479	0.107363	-0.08136	0.9353	
oslo	0.0458516	0.122697	0.3737	0.7093	
Mean dependent var	1.927636	S.D. dependent var	1.151154		
Sum squared resid	12.84528	S.E. of regression	0.324483		
R-squared	0.923674	Adjusted R-squared	0.920546		
F(5, 122)	295.2809	P-value(F)	2.29e-66		

f) Tolk utskriften.

En medievider har påstått at redaksjonelle kostnader øker «eksponensielt» med avisenes opplag og at det å erstatte momsfrirket med støtte basert på skattefradrag for redaksjonelle kostnader derfor vil føre til en omfordeling av støtte fra små til store aviser. Under finner du et plott av avisenes redaksjonelle kostnader mot deres opplag. En regresjonslinje basert på et fjerdeordenspolynom i opplaget er også tegnet inn.



g) Er medieviderens kritikk relevant?

Figuren over gir grunn til å tro at feilleddet er heteroskedastisk. Gå tilbake til modellen $Redkost = \alpha + \beta \cdot opplag + \varepsilon$. Anta at standardavviket til feilleddet ε øker med avisenes opplag slik at

$$S(\varepsilon) = opplag \cdot \sigma$$

der σ er en konstant.

b) Vis at følgende transformerte modell har homoskedastisk feilledd:

$$Redkost/opplag = \alpha/opplag + \beta + \varepsilon/opplag$$

Du skal altså vise at $\text{Var}(\varepsilon/opplag)$ er konstant. Du kan betrakte avisenes opplag som en deterministisk (ikke-stokastisk) variabel.

Oppgave 3

I kurset INT010 ble det holdt 18 forelesninger hvor det gjennomsnittlig ble gjennomgått Y_t powerpointplanser per forelesning. Under ser du en graf med antall planer gjennomgått på hver av de 18 forelesningene.



Anta at forelesningsrekken fortsatte i de uendelige!

- Drøft kort om «hvit støy» eller «tilfeldig gang» vil være den beste tilnærmingen til prosessen dersom du kun kan velge mellom disse to.
- Drøft kort om en prosess med positiv eller negativ autokorrelasjon kan være aktuell i denne konteksten. Foreslå en enkel modell som åpner for autokorrelasjon i antall planer per forelesning.