

NORGES HANDELSHØYSKOLE

Eksamen høstsemesteret 2010

Kurskode: INT010 Tittel: Anvendt metode

Dato: 05.11.2010 Kl. 09.00-12.00

Faglærer går ikke rundt i eksamenslokalene, men kan kontaktes av eksamensvakten på tlf. 59670/41645914.

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

Vi ønsker å lage en enkel modell for å predikere etterspørsel for bensin i USA. Fra økonomisk teori vet vi at etterspørsel avhenger av inntekt og pris. Årlige data er samlet inn for perioden t=1960-1995 (dvs. 36 årlige observasjoner):

- Årlig bensinforbruk
- En prisindeks for bensin (normalisert til 1 i 1967)
- Inntekt per capita

Vi analyserer først en regresjonsmodell med logaritmisk årlig bensinforbruk (logGpc) som responsvariabel, og de følgende forklaringsvariable; logaritmisk prisindeks for bensin (logPG), logaritmisk inntekt per capita (logI), år (YR) og år kvadrert (YRSQ). Minitab gir følgende utskrift:

Regression Analysis: logGpc versus logPG; logI; YR; YRSQ

The regression equation is

$$\text{logGpc} = -3,92 - 0,160 \text{ logPG} + 1,26 \text{ logI} - 0,00059 \text{ YR} - 0,000129 \text{ YRSQ}$$

Predictor	Coef	SE Coef	T	P
Constant	-3,925	1,479	-2,65	
logPG	-0,16016	0,03390	-4,72	
logI	1,2612	0,3746	3,37	
YR	-0,000594	0,004375	-0,14	
YRSQ	-0,00012900	0,00007560	-1,71	

S = 0,0122644 R-Sq = 96,9% R-Sq(adj) = 96,5%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	0,145413	0,036353	241,68	
Residual Error	31	0,004663	0,000150		
Total	35	0,150076			

Source	DF	Seq SS
logPG	1	0,066160
logI	1	0,070893
YR	1	0,007922
YRSQ	1	0,000438

Unusual Observations

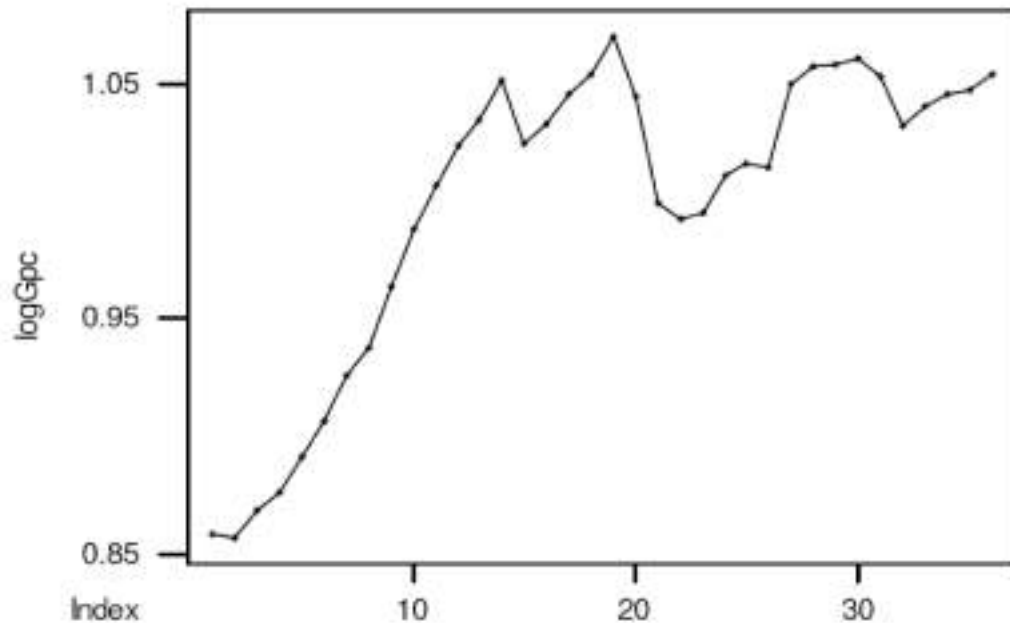
Obs	logPG	logGpc	Fit	SE Fit	Residual	St Resid
1	-0,034	0,85772	0,83505	0,00791	0,02267	2,42R
8	0,000	0,93752	0,96252	0,00327	-0,02500	-2,11R

R denotes an observation with a large standardized residual.

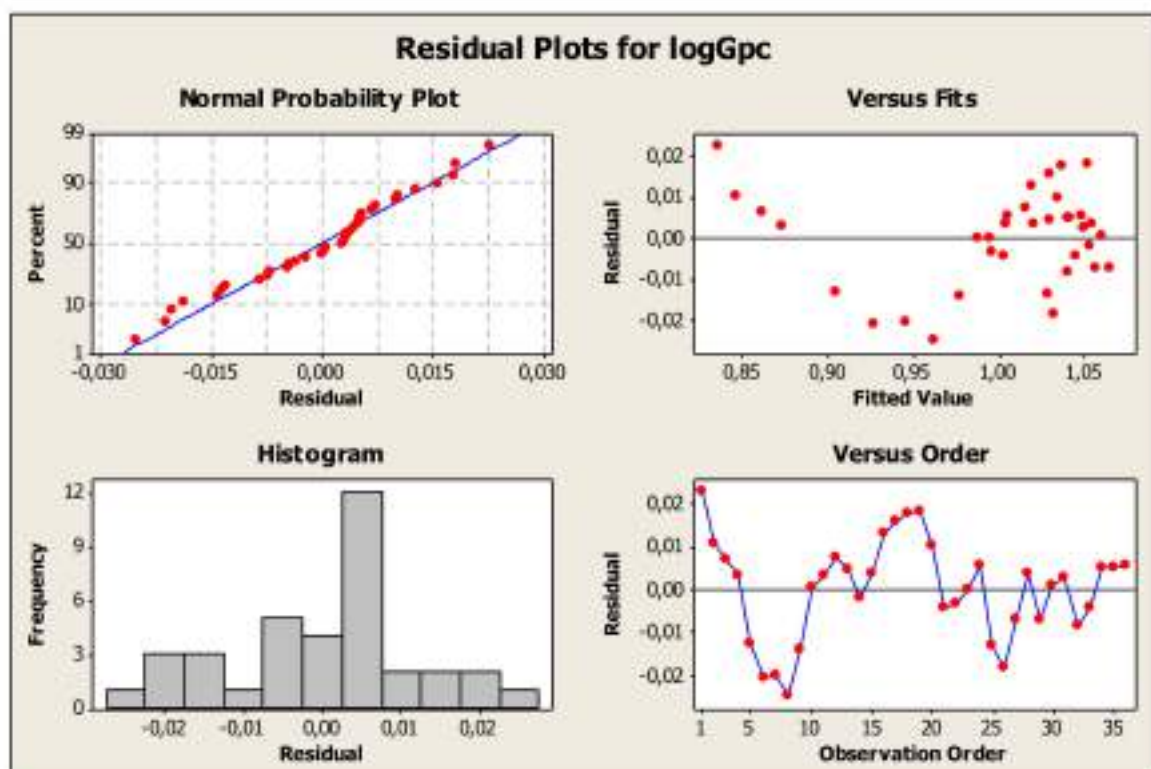
Durbin-Watson statistic = 0,512145

- Hvilke koeffisienter er signifikant forskjellig fra null på 5%-nivå? Angi kritisk verdi og antall frihetsgrader på testobservatoren.
- Beregn et 95% konfidensintervall for koeffisienten til variabelen logaritmisk prisindeks for bensin (logPG).

- c) I figuren under er logGpc (dvs. logaritmisk bensinforbruk i hvert år) plottet mot år. Kan du forklare med bakgrunn i denne figuren hvorfor det gir mening å inkludere år og år kvadrert (dvs. år^2) som forklaringsvariabler i regresjonen?



- d) Drøft kort resultatene fra regresjonen. Vurder spesielt om fortegnene til koeffisientene er rimelige.
- e) Det er gjort noen analyser av residualene, se under. Forklar hvilken informasjon de ulike plottene gir.



- f) Vis at feilleddene i regresjonsanalysen er positivt autokorrelert av første orden ved å gjennomfør en test på 5% signifikansnivå. Angi forkastningsområde for testobservatoren.
- g) Dersom feilleddene i en regresjon er autokorrelerte, hvilke konsekvenser vil det ha for analysen?

Det er mulig å korrigere for autokorrelasjon i feilleddene ved å benytte en lagget versjon av responsvariabelen som forklaringsvariabel, dvs. at forrige års konsum av bensin kan være med og forklare dette års konsum (dvs. om responsen er y_t er den laggete variabelen y_{t-1}). Videre kan det være rimelig å inkludere en lagget versjon av prisindeksen til bensin. I regresjonen under er derfor to nye forklaringsvariabler inkludert; forrige års bensinkonsum (Lagged logGpc) og forrige års prisindeks (Lagged logPG). Det kan også være rimelig å fjerne YR, den minst signifikante av de to tidsvariablene, fra regresjonen. Utskrift og diagnoseplott fra Minitab er gitt under:

Regression Analysis: logGpc versus Lagged logGpc; logPG; Lagged logPG; logI; YRSQ

The regression equation is:

$$\text{logGpc} = -1,47 + 0,704 \text{ Lagged logGpc} - 0,271 \text{ logPG} + 0,183 \text{ Lagged logPG} + 0,457 \text{ logI} - 0,000059 \text{ YRSQ}$$

35 cases used, 1 cases contain missing values

Predictor	Coef	SE Coef	T	P
Constant	-1,4723	0,3343	-4,40	0,000
Lagged logGpc	0,70380	0,08982	7,84	0,000
logPG	-0,27085	0,03250	-8,33	0,000
Lagged logPG	0,18266	0,03904	4,68	0,000
logI	0,4571	0,1070	4,27	0,000
YRSQ	-0,00005874	0,00001541	-3,81	0,001

S = 0,00652111 R-Sq = 99,0% R-Sq(adj) = 98,9%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	5	0,128113	0,025623	602,53	0,000
Residual Error	29	0,001233	0,000043		
Total	34	0,129347			

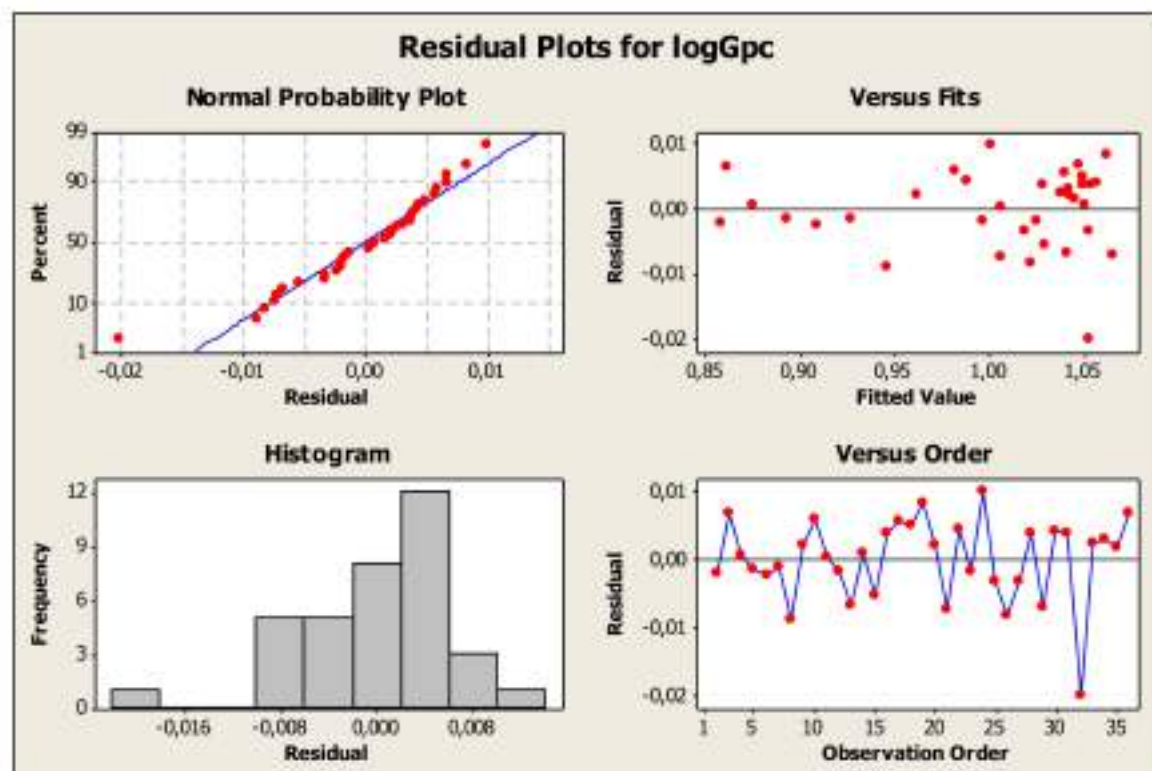
Source	DF	Seq SS
Lagged logGpc	1	0,121782
logPG	1	0,000529
Lagged logPG	1	0,004972
logI	1	0,000212
YRSQ	1	0,000618

Unusual Observations

Obs	Lagged logGpc	logGpc	Fit	SE Fit	Residual	St Resid
32	1,05	1,03261	1,05271	0,00254	-0,02009	-3,35R

R denotes an observation with a large standardized residual.

Durbin-Watson statistic = 2,14755



h) Drøft kort resultatene fra den nye regresjonsanalysen.

Oppgave 2

En bilprodusent hevder at ikke mer enn 10% av bilene behøver reparasjoner i løpet av de første tre årene etter bilen er produsert (garantiperioden). I ett tilfeldig utvalg av 50 tre år gamle biler var det 8 biler som hadde blitt reparert. Vi er interessert i å finne ut om dette medfører at bilprodusentens påstand er usann.

- Hvilken hypotesetest kan benyttes i denne situasjonen? Sett opp testobservatoren. Hvilken fordeling har testobservatoren?
- Formuler null og alternativhypotesen, og gjennomfør testen på 5% signifikansnivå. Kommenter.
- Finn p-verdien tilhørende testen i b).

Oppgave 3

En produsent av ett kjent frokostblandingsprodukt ønsker å profilere det høye rosinnholdet i frokostblandingen. Produsenten garanterer at hver porsjon (av en bestemt størrelse) en person spiser av frokostblandingen skal inneholde minst *en* rosin med en (høy) sannsynlighet β . En

pakke med frokostblanding har innhold til m porsjoner og inneholder n rosiner som vi antar er uniformfordelt i pakken. La X_1, X_2, \dots, X_m være det antall rosiner en person får i porsjon nummer $1, 2, \dots, m$.

- Hva er fordelingen til X_1 ? Begrunn svaret.
Hva er fordelingen til X_2 gitt at $X_1 = x_1$?
- Finn et uttrykk for hvor mange rosiner n det må minst tilsettes for at det er minst en sannsynlighet på β for at den første porsjonen skal inneholde minst en rosin?
- Er X_1 og X_2 uavhengige? Begrunn svaret.
Hva er fortegnet på korrelasjonskoeffisienten mellom X_1 og X_2 ? Begrunn svaret.

Produsenten ønsker å undersøke hvor mange av de m porsjonene av frokostblandingen som er uten rosiner. La W betegne dette antallet. For å forenkle beregninger uttrykker vi W ved hjelp av tellevariable Y_1, Y_2, \dots, Y_m ,

$$W = \sum_{i=1}^m Y_i, \quad \text{der} \quad Y_i = \begin{cases} 1, & X_i = 0 \\ 0, & X_i > 0 \end{cases}$$

- Finn $E(Y_1)$ og $E(W)$.