

NORGES HANDELSHØYSKOLE

Eksamen vårsemesteret 2011

Kurskode: INT010 Tittel: Anvendt metode

Dato: 24.05.2011 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

På et planlagt lederseminar er det deltagere fra ulike næringer; industri (1), finans (2) og ingeniørbedrifter (3). Før seminaret ønsker arrangøren å kartlegge om det er forskjeller i kunnskapen blant deltakerne fra de tre ulike næringene. En test for å kartlegge kunnskapsnivået er blitt gjennomført og hver deltaker har fått en score mellom 1 og 110. Gjennomsnittscore for de ulike gruppene er hhv. $\bar{X}_1 = 52$, $\bar{X}_2 = 80$ og $\bar{X}_3 = 57$ og de tilhørende standardavvikene er $S_1 = 13$, $S_2 = 24$ og $S_3 = 22$. Antall deltakere fra hver næring er $n_1 = 7$, $n_2 = 8$ og $n_3 = 6$.

- Vis at hypotesen om like varianser i næring 1 (industri) og 2 (finans) blir akseptert på 5% signifikansnivå.
- Formuler nullhypotesen for å undersøke om forventet kunnskapsnivå er likt mellom deltakerne fra industri og finans, mot alternativet at de er ulike. Sett opp testobservator og kritisk verdi for testobservatoren på 5 % signifikansnivå når vi antar at standardavviket til de to næringene er like.
- Hvilke forutsetninger må oppfylles for å benytte testen i b)? Hva er konklusjonen på testen i b) når du benytter Minitab-utskriften under.

Two-Sample T-Test and CI: Industri; Finans

Two-sample T for Industri vs Finans

	N	Mean	StDev	SE Mean
Industri	7	52,3	12,5	4,7
Finans	8	80,3	24,0	8,5

Difference = mu (Industri) - mu (Finans)

Estimate for difference: -28,0

95% CI for difference: (-49,9; -6,1)

T-Test of difference = 0 (vs not =): T-Value = -2,76 P-Value = 0,016 DF = 13

Both use Pooled StDev = 19,5781

For å sammenligne kunnskapsnivået mellom alle de tre næringene, gjennomføres det en en-veis variansanalyse. Under finner du en utskrift fra Minitab.

One-way ANOVA: Kunnskap versus Næring

Source	DF	SS	MS	F	P
Næring	2	3395	1697	4,12	0,034
Error	18	7416	412		
Total	20	10811			

S = 20,30 R-Sq = 31,40% R-Sq(adj) = 23,78%

- Drøft kort resultatet fra den utførte ANOVA.

Det er også gjennomført en Kruskal-Wallis test. Se utskrift fra Minitab under.

Kruskal-Wallis Test: Kunnskap versus Næring

Kruskal-Wallis Test on Kunnskap

Næring	N	Median	Ave Rank	Z
1	7	50,00	8,2	-1,45
2	8	88,00	15,1	2,39
3	6	51,50	8,8	-1,05
Overall	21		11,0	

H = 5,74 DF = 2 P = 0,057

H = 5,75 DF = 2 P = 0,057 (adjusted for ties)

- e) Formuler null og alternativhypotesen for Kruskal-Wallis testen. Hva er forkastningsområdet til testobservatoren på 5% signifikansnivå? Hva blir konklusjonen på testen? Sammenlign med svaret i d) og kommenter.

Oppgave 2

En økonom ønsker å analysere sammenhengen mellom lønn (Wage), utdanning (Edu) og arbeidserfaring (Exper). Data over 523 arbeidstagere i USA er samlet inn. Lønn er uttrykt i US\$ per time, utdanning er antall år på skole/universitet, og arbeidserfaring er antall år i jobb.

Minitab gir følgende utskrift for en regresjon med lønn som respons og to forklaringsvariable:

Regression Analysis: Wage versus Educ; Exper

The regression equation is

Wage = - 4,52 + 0,913 Educ + 0,0968 Exper

Predictor	Coef	SE Coef	T	P
Constant	-4,524	1,239	-3,65	0,000
Educ	0,91302	0,08219	11,11	0,000
Exper	0,09681	0,01772	5,46	0,000

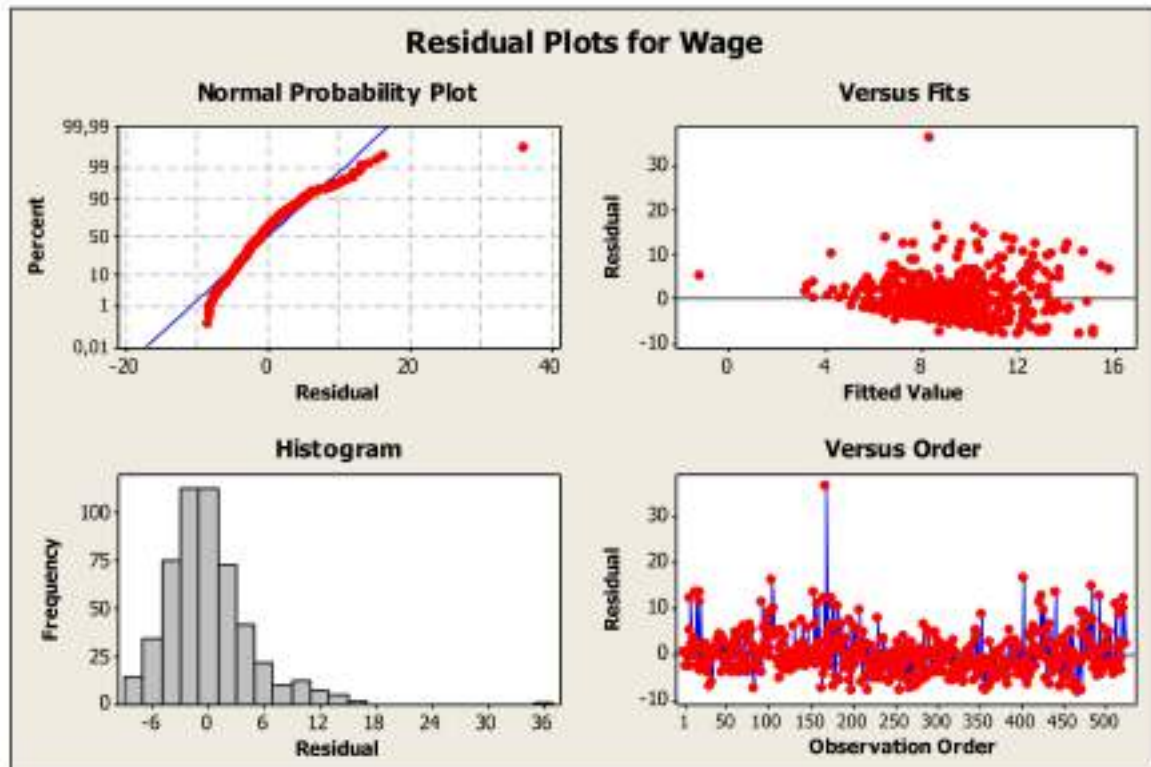
S = 4,62357 R-Sq = 19,5% R-Sq(adj) = 19,2%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	2691,9	1346,0		
Residual Error	520	11116,3	21,4		
Total	522	13808,2			

Source	DF	Seq SS
Educ	1	2053,8
Exper	1	638,1

- Test hypotesen H_0 : Alle koeffisienter i modellen (unntatt konstantleddet) er lik null mot H_1 : Minst en forklaringsvariabel (utenom konstantleddet) har koeffisient ulik null.
- Beregn et 95% konfidensintervall for koeffisienten til variabelen utdanning (Educ).
- Drøft kort resultatene fra regresjonen. Vurder spesielt om fortegnene til koeffisientene er rimelige.
- Det er gjort noen analyser av residualene, se under. Forklar hvilken informasjon de ulike plottene gir. Er modellantagelsene oppfylt?



For å teste formelt om det er problemer med heteroskedastisitet i regresjonsanalyser, kan man gjennomføre **White's test for heteroskedastisitet**, hvor nullhypotesen er ingen heteroskedastisitet.

Prosedyren for å gjennomføre denne testen er som følger: anta vi har en regresjonsmodell med to forklaringsvariable, dvs. $Y_i = \beta_0 + \beta_1 \cdot X_{i1} + \beta_2 \cdot X_{i2} + e_i$, med totalt n antall observasjoner. Først estimeres denne regresjonsmodellen, og residualene (e_i) blir estimert. I steg to estimeres den følgende regresjonsmodell (vi kaller den B),

$$e_i^2 = \alpha_0 + \alpha_1 \cdot X_{i1} + \alpha_2 \cdot X_{i2} + \alpha_3 \cdot X_{i1}^2 + \alpha_4 \cdot X_{i2}^2 + \alpha_5 \cdot X_{i1} \cdot X_{i2} + u_i$$

Fra denne regresjonen får vi en forklaringsgrad (R^2). Testobservatoren (W) for White's test er da:

$$W = n \cdot R^2 \quad (\text{merk at } R^2 \text{ ikke er i } \%, \text{ men tall. F.eks. ikke } 10\%, \text{ men } 0,1)$$

Det kan vises at denne testobservatoren er kjikvadrat-fordelt med k frihetsgrader, der k er antall forklaringsvariable i regresjonsmodellen for e_i^2 over. Nullhypotesen blir altså forkastet om $W > \chi^2_{\alpha, k}$ der α er signifikansnivået.

Fra regresjonen for lønn over, har vi tatt vare på residualene (e_i), og estimerer regresjonsmodellen B. Resultatene fra Minitab er som følger:

Regression Analysis: Residual^2 versus Educ; Exper; Educ^2; Exper^2; Educ*Exper

The regression equation is

$$\text{Residual}^2 = 14,4 - 1,18 \text{ Educ} - 1,40 \text{ Exper} + 0,169 \text{ Educ}^2 + 0,0271 \text{ Exper}^2 + 0,022 \text{ Educ*Exper}$$

Predictor	Coef	SE Coef	T	P
Constant	14,38	71,35	0,20	0,840
Educ	-1,183	9,138	-0,13	0,897
Exper	-1,401	1,912	-0,73	0,464
Educ^2	0,1686	0,3007	0,56	0,575
Exper^2	0,02711	0,02097	1,29	0,197
Educ*Exper	0,0222	0,1041	0,21	0,831

S = 65,1437 R-Sq = 2,15% R-Sq(adj) = 1,20%

- c) Benytt informasjon fra Minitab utskriften over for å regne ut verdien på testobservatoren (W) for White's test for heteroskedastisitet. Hva er kritisk verdi for testobservatoren på 5% signifikansnivå? Hva er konklusjonen på testen?

Økonomen bestemmer seg så for å transformere responsen lønn ved en logaritmisk transformasjon, dvs. $\ln(\text{lønn})$. Denne transformerte responsen er så benyttet i en regresjon med utdanning og arbeidserfaring som forklaringsvariable. Se utskrift fra Minitab under.

Regression Analysis: logWage versus Educ; Exper

The regression equation is

$$\text{logWage} = 0,656 + 0,0942 \text{ Educ} + 0,0105 \text{ Exper}$$

Predictor	Coef	SE Coef	T	P
Constant	0,6564	0,1253	5,24	0,000
Educ	0,094152	0,008312	11,33	0,000
Exper	0,010477	0,001792	5,85	0,000

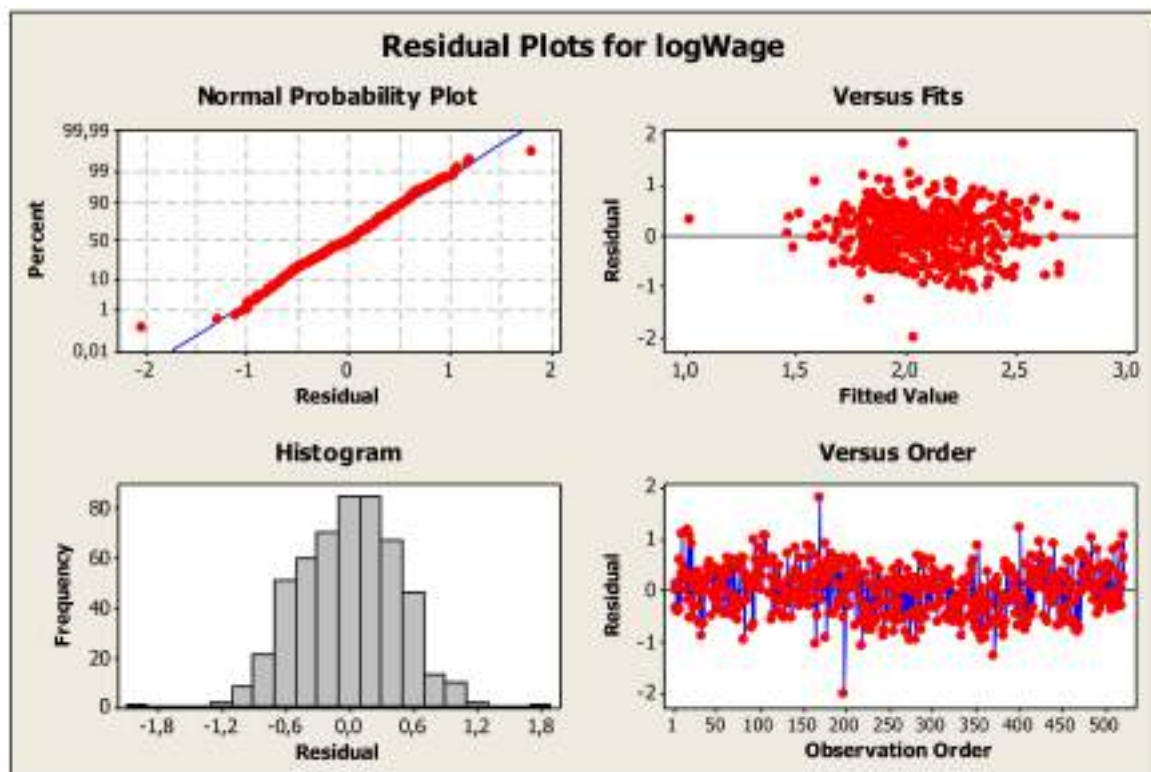
S = 0,467563 R-Sq = 20,2% R-Sq(adj) = 19,9%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	28,854	14,427	65,99	0,000
Residual Error	520	113,680	0,219		
Total	522	142,534			

Source	DF	Seq SS
Educ	1	21,381
Exper	1	7,473

- f) Drøft kort resultatene fra regresjonen med $\ln(\text{lønn})$ som respons. Tolk spesielt koeffisientene til de to forklaringsvariablene.
- g) Under vises diagnoseplott av residualene fra regresjonen i f). Kommenter kort plottene og sammenlign med plottene fra d).



Oppgave 3

Prisen P_n på en aksje observeres i etterfølgende tidspunkter $n = 0, 1, 2, 3, \dots$. Prisen ved starttidspunktet, P_0 , er gitt og i hver etterfølgende periode kan prisen enten gå opp a prosent eller ned b prosent fra den foregående. La

$$I_n = \begin{cases} 1 & \text{dersom oppgang i periode } n \\ 0 & \text{dersom nedgang i periode } n \end{cases}$$

- a) Forklar at vi kan skrive

$$P_n = P_{n-1}(1+a)^{I_n}(1-b)^{1-I_n} = P_0(1+a)^{X_n}(1-b)^{n-X_n} = P_0(1-b)^n \left(\frac{1+a}{1-b} \right)^{X_n}$$

der X_n = antall oppganger i de n periodene.

- b) Anta at i hver periode har oppgang/nedgang sannsynligheter hhv. p og $1-p$, og at utfallene i de enkelte perioder er uavhengige. Vis at (se Hint på neste side)

$$EP_n = P_0 \cdot (1 - b(1-p) + ap)^n$$

c) Beregn EP_n for $a = 0,20$, $b = 0,10$ og $p = 0,4$ i tilfellene $n = 2$ og $n = 10$.

Hint for b): Bruk at når X er binomisk fordelt med suksessannsynlighet p og n uavhengige observasjoner så er

$$Ez^X = (1 - p + pz)^n \quad \text{for alle reelle tall } z$$