

NORGES HANDELSHØYSKOLE

Eksamen høst 2012

Kurskode: INT010 Tittel: Anvendt metode

Dato: 27.11.2012 Kl. 09.00-12.00

Faglærer går ikke rundt i eksamenslokalene, men kan kontaktes av eksamensvakten på
tlf. 59248/924 27 289

Hjelpemidler til eksamen:

Hjelpemidler tillatt: JA, alle trykte/egenskrevne

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

Alle delspørsmål i oppgavesettet teller likt.

Oppgave 1

Studenter i bachelorstudiet på NHH har rett til å fortsette på master etter avlagt bachelorgrad. Dette omtales som det «interne opptaket» til masterstudiet. I tillegg kan bachelorstudenter fra andre institusjoner søke seg til masterstudiet på NHH. Dette omtales som et «eksternt opptak». I det eksterne opptaket er det en opptakskvote for studenter med grad fra Norden, og en annen opptakskvote for internasjonale studenter fra ikke-nordiske land. Da det er stor heterogenitet i utdanningssystemer utenfor Norden – både hva gjelder fagkombinasjoner og karakterskalaer – er det en større grad av forskjell ved opptaket av ikke-nordiske studenter.

Det er delte meninger i staben på NHH om de eksternt opptatte studentene klarer seg bedre eller dårligere i masterstudiet enn de interne. Videre er det ønskelig å balansere den faglige kvaliteten best mulig mellom de to eksterne gruppene (nordiske og fra øvrige land) slik at det ikke blir merkbar forskjell på de to gruppene.

Dekanen for masterstudiet ønsker å kvalitetssikre størrelsen på det eksterne opptaket. Han ber studieadministrasjonen lage et datasett med observasjoner for hver masterstudent tatt opp i årene 2004 til 2010 og som har fullført studiet. Følgende variabler er inkludert:

snittkar_master – gjennomsnittskaracteren i masterstudiet for hver student. For å kunne regne ut gjennomsnittet er bokstavkarakterer gjort om til tallkarakterer: A=5, B=4 osv.

opptak – en variabel som viser om studentene er tatt opp «internt» eller «eksternt».

opptak2 – en variabel som viser om studentene er tatt opp «internt», «eksternt nordisk» eller «internasjonalt» (internasjonale er eksterne ikke-nordiske studenter).

forsinket – dummyvariabel med verdi én dersom studenten brukte mer enn 4 semestre, null ellers.

poengpersem – gjennomsnittlig antall studiepoeng tatt per semester.

Med utgangspunkt i dette datasettet blir det gjort følgende analyser:

ANALYSE 1:

Two-Sample T-Test and CI: snittkar_master; opptak

Two-sample T for snittkar_master

opptak	N	Mean	StDev	SE Mean
Eksternt	746	3,837	0,472	0,017
Internt	1147	4,023	0,494	0,015

Difference = mu (Eksternt) - mu (Internt)
Estimate for difference: -0,1858
95% CI for difference: (-0,2306; -0,1410)
T-Test of difference = 0 (vs not =): T-Value = -8,13 P-Value = 0,000 DF = 1891
Both use Pooled StDev = 0,4856

ANALYSE 2:

One-way ANOVA: snittkar_master versus opptak2

Source	DF	SS	MS	F	P
opptak2	2	15,639	7,819	33,14	0,000
Error	1890	445,933	0,236		
Total	1892	461,571			

S = 0,4857 R-Sq = 3,39% R-Sq(adj) = 3,29%

				Individual 95% CIs For Mean Based on Pooled StDev			
Level	N	Mean	StDev	-+-----+-----+-----+-----			
Eksternt nordisk	570	3,8410	0,4419	(----*----)			
Internasjonalt	176	3,8253	0,5592	(-----*-----)			
Internt	1147	4,0231	0,4944	(----*--)			
				-+-----+-----+-----+-----			
				3,760	3,840	3,920	4,000

Pooled StDev = 0,4857

ANALYSE 3

Two-Sample T-Test and CI: poengpersem; opptak

Two-sample T for poengpersem

opptak	N	Mean	StDev	SE Mean
Eksternt	739	27,56	6,51	0,24
Internt	1114	27,84	7,99	0,24

Difference = mu (Eksternt) - mu (Internt)

Estimate for difference: -0,281

95% CI for difference: (-0,945; 0,383)

T-Test of difference = 0 (vs not =): T-Value = -0,83 P-Value = 0,406 DF = 1774

ANALYSE 4:

Test and CI for Two Proportions: forsinket; opptak

Event = 1

opptak	X	N	Sample p
Eksternt	235	739	0,317997
Internt	335	1114	0,300718

Difference = p (Eksternt) - p (Internt)

Estimate for difference: 0,0172792

95% CI for difference: (-0,0257615; 0,0603198)

Test for difference = 0 (vs not = 0): Z = 0,79 P-Value = 0,431

Legg 5% signifikansnivå til grunn i din besvarelse av spørsmålene under.

- a) Kommenter kort analyse 1 og 2.
- b) Kommenter kort analyse 3 og 4.
- c) Gi en kort innstilling til dekanen basert på analysene.
- d) Dekanen kommer tilbake til deg og lurere på om det er signifikant forskjell mellom de to gruppene av eksterne studenter med hensyn til variansen til gjennomsnittskarakterene deres. Undersøk dette.
- e) Dekanen ber deg også teste om andelen som slutter er den samme for interne og eksterne studenter. Han gir deg følgende tabell der raden med sluttet=1 gir deg antall studenter som har sluttet i masterstudiet uten å fullføre: (Utvalget er nå utvidet slik at det inkluderer alle studenter som ikke har fullført.)

sluttet	Opptak		Total
	Eksternt	Internt	
0	825	1,240	2,065
1	68	95	163
Total	893	1,335	2,228

- f) Dekanen påstår at analysene dine bygger på en antagelse om normalfordelte variabler.
 - (i) Han spør om karakterdataene i analysen kan betraktes som normalfordelte
 - (ii) Uavhengig av ditt svar på (i) insisterer han på å få se robusthetsanalyser som ikke bygger på normalitet. Hvilke analyser vil du anbefale?

Oppgave 2

Universitets- og høyskolerådet følger karaktersettingen på medlemsinstitusjonene. Det er reist spørsmål ved om NHH bruker karakterskalaen feil og gir for gode karakterer på masteroppgavene. Dekanen har bestilt en regresjonsanalyse for å prøve og forstå karaktersettingen bedre. Han er spesielt opptatt eventuelle forskjeller i karaktersetting på «enkle» og «doble» utredninger. Doble utredninger er utredninger som er skrevet av to studenter i felleskap. På NHHs nettsider for studenter finner man følgende informasjon om masteroppgaven:

Arbeidet kan gjennomføres individuelt eller sammen med en medstudent. ... Det tillates ikke at mer enn to studenter skriver sammen Mange erfarer at arbeidet blir mer lærerikt og givende hvis man er flere som skriver sammen. ... Ved samarbeid om oppgaven skal omfanget tilsvare summen av to studenters innsats.... En kollektiv utredning innleveres til sensur under ett for gruppens medlemmer. Vedkommende veileder vil skriftlig erklære at han/hun er forvisset om at begge medlemmer av gruppen har bidratt til arbeidet med utredningen i rimelig grad

Den avhengige variabelen er tallkarakteren på masterutredninger med A=5, B=4, C=3 osv. A skal representere en «fremragende prestasjon som klart utmerker seg». C skal være en «jevnt god prestasjon» osv. E «tilfredsstiller minimumskravene, men heller ikke mer». Datasettet inneholder alle masterutredninger levert i årene 2006-2011. Variablene er definert som følger:

Gjennomsnittskarakter er gjennomsnittet av alle karakterene tidligere i masterstudiet for den eller de studentene som har skrevet oppgaven.

Enkel er en dummyvariabel med verdi 1 dersom utredningen er skrevet av én student og null dersom utredningen er skrevet av to studenter.

Dobbel er en dummyvariabel med verdi 1 dersom utredningen er skrevet av to studenter og null dersom utredningen er skrevet av én student.

Max_gjennomsnittskarakter er gjennomsnittskarakteren til den forfatteren av en utredning med høyest karaktersnitt

Min_gjennomsnittskarakter er gjennomsnittskarakteren til den forfatteren av en utredning med lavest karaktersnitt

Karakterdifferanse er ($\text{Max_gjennomsnittskarakter} - \text{Min_gjennomsnittskarakter}$), altså karakterforskjellen mellom den forfatteren med høyest og lavest gjennomsnittskarakter. Dersom det bare er én forfatter blir variabelen null.

Tabell 1: Deskriptiv statistikk

	Antall obs.	Gjennomsnitt	Standard- avvik	Min. verdi	Max. Verdi
Karakter på masteroppgaven, enkle utredninger	1001	4,160	0,839	0	5
Karakter på masteroppgaven, doble utredninger	479	4,482	0,592	2	5
Gjennomsnittskarakter tidligere i masterstudiet, enkle utredninger	1001	3,720	0,577	1,658	5
Gjennomsnittskarakter tidligere i masterstudiet, doble utredninger	479	3,790	0,442	2,362	4,813
Karakterdifferanse, doble utredninger	479	0,460	0,376	0	2,283

Tabell 2: Regresjonsanalyse av karaktersetting på masterutredninger

	(1)	(2)	(3)	(4)	(5)
Konstantledd	2,101*** (0,130)	1,957*** (0,146)	1,957*** (0,146)	1,957*** (0,146)	1,986*** (0,150)
Gjennomsnittskarakter	0,554*** (0,0343)				
Dobbel	0,282*** (0,0394)	0,949*** (0,315)	0,883*** (0,336)	0,883*** (0,336)	0,924*** (0,336)
Gjennomsnittskarakter*Enkel		0,592*** (0,0388)	0,592*** (0,0388)	0,592*** (0,0388)	0,582*** (0,0392)
Gjennomsnittskarakter*Dobbel		0,416*** (0,0732)	0,427*** (0,0759)		
Karakterdifferanse			0,0506 (0,0891)		
Max_gjennomsnittskarakter*Dobbel				0,264** (0,106)	0,242** (0,107)
Min_gjennomsnittskarakter* Dobbel				0,163* (0,0871)	0,162* (0,0878)
<i>N</i>	1480	1480	1480	1480	1480
Justert R^2	0,180	0,182	0,182	0,182	0,217
MSE	0,7	0,7	0,7	0,7	0,7

Koeffisientenes standardavvik er oppgitt i parentes. Spesifikasjonen i kolonne (5) inneholder 65 dummier for veiledere. Koeffisientene til disse dummier er ikke rapportert i tabellen.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

- Drøft kort hva vi lærer fra den deskriptive analysen i tabell 1.
- Tolk koeffisienten til dummyvariabelen «Dobbel» i kolonne (1) og presenter minst to mulige forklaringer på hvorfor den er signifikant positiv.
- Bruk modellene i kolonne (3) og (4) til å predikere forventet karakter for to utredninger der
 - To studenter skal skrive sammen og begge har et karaktersnitt på 4
 - To studenter skal skrive sammen. Den beste har et karaktersnitt på 4,5 og den dårligste har et karaktersnitt på 3,5.
- Bruk kolonne (3) og (4) til å drøfte om det foreligger et «gratispassasjerproblem» når to studenter skriver sammen, det vil si om vi kan forkaste en hypotese om at studentene bidrar likt til karakteren på en felles utredning.
- Representerer kolonne (3) og (4) to forskjellige regresjonslinjer (hyperplan), eller er de helt sammenfallende? Forklar.

Datasettet inneholder informasjon om hvem som har veiledet de ulike oppgavene. Hver veileder har et id-nummer. Anta for eksempel at professor Jarle Møen har veileder-id lik 23. Da finnes det en dummyvariabel, D23 som har verdien 1 dersom en oppgaven ble veiledet av Jarle Møen og null ellers. Spesifikasjonen i kolonne (5) inneholder 65 slike veilederdummier. Alle veiledere som har veiledet minst ti oppgaver er representert med egen dummy. Av de 65 veilederdummiene er 11 signifikante på 5%-nivå. Den største koeffisientverdien blant disse 11 er 0,491 for D30, og den minste er -1,067 for D10.

- f) Hva er tolkningen av koeffisientene til veilederdummiene? Ta utgangspunkt i den oppgitte min- og maksverdien blant de signifikante koeffisientene.
- g) Hvor mange signifikante veilederdummier (f.eks. på 5% nivå) vil du forvente å finne dersom alle veiledere er like dyktige og bruker karakterskalaen likt?

Oppgave 3

En långiver har et utestående lån til en låntaker som kan gå konkurs eller forbli solvent. Tapet T kan da skrives $T=Y \cdot I$, der Y er tapet ved en konkurs, og $I=1$ ved konkurs og 0 ellers. Anta at sannsynligheten for konkurs er p, og at Y og I er uavhengige (uten tap av generalitet).

a) Vis at

$$(i) \quad ET = p \cdot EY$$

$$(ii) \quad \text{var}(T) = p \cdot \text{var}(Y) + p(1-p) \cdot (E(Y))^2$$

Hint: Husk at generelt gjelder $\text{var}(X) = E(X^2) - (E(X))^2$

En enkel modell for sannsynligheten for konkurs basert på en observerbar indikator Z er

$$D(z) = P(I = 1 | Z = z) = \frac{e^{\alpha + \beta z}}{1 + e^{\alpha + \beta z}} = \frac{1}{1 + e^{-(\alpha + \beta z)}}$$

- b) Forklar kort at dette gir mening som modell, dvs. at $D(z)$ representerer sannsynligheter med fornuftige egenskaper.
- c) Vis at log-odds er lineær i z, dvs at $\ln(P(I = 1 | Z = z) / P(I = 0 | Z = z)) = \alpha + \beta z$.