

## NORGES HANDELSHØYSKOLE

### Eksamen høst 2014

Kurskode: INT010      Tittel: Anvendt metode

Dato: 28.10.2014      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.  
Husk at dere må skrive slik at sensor kan lese det som er skrevet.***

*Eksamensoppgaven består av 7 sider, inkludert denne siden.*

### Oppgave 1

Ved eksamen i anvendt metode våren 2007 var det 370 oppmeldte kandidater. Av disse møtte 347 til eksamen. De ble fordelt på seks sensurkommisjoner. Alle besvarelser ble rettet av én førstesensor og én andresensor som hver for seg ga besvarelsene poeng på en skala fra 1 til 100. Tabellen under viser antall besvarelser som ble rettet av hver av de seks kommisjonene, gjennomsnittlig poeng gitt av hhv første- og andresensor og standardavvikene til poengene. Kandidatnumrene var tildelt av eksamenskontoret i tilfeldig rekkefølge. Dataene er reelle, men sensornavnene er fiktive.

Variabelen *Poeng1* er poeng gitt av førstesensor og variabelen *Poeng2* er poeng gitt av andresensor. Differanse er (*Poeng1*-*Poeng2*).

Kommisjon	Antall Kandidater	Første- sensor	Gj.snitt Poeng1	St.avvik Poeng1	Andre- sensor	Gj.snitt Poeng2	St.avvik Poeng2	Gj.snitt Differanse	St.avvik Differanse
1	59	Bårdsen	68.06	19.03	Nansen	75.10	18.80	-7.04	6.44
2	59	Bårdsen	64.05	17.19	Trondsen	66.07	17.11	-2.01	5.13
3	58	Bårdsen	69.62	16.31	Kristensen	68.28	14.49	1.34	6.52
4	59	Eriksen	66.02	16.84	Nansen	68.69	19.51	-2.67	5.41
5	55	Eriksen	65.72	17.23	Trondsen	65.25	18.12	0.46	4.18
6	57	Eriksen	69.98	15.00	Kristensen	66.18	15.42	3.81	4.80
Totalt	347		67.24	17.00		68.31	17.53	-1.07	6.46

Førstesensor Bårdsen har lite sensorerfaring. Vi lurte på om hans rettestil er den samme i den første bunken han retter som i den siste.

- Test om variansen i førstesensor Bårdsens poenggivning er den samme i kommisjon 1 som i kommisjon 3.
- Test om forventet poengsum gitt av førstesensor Bårdsen i kommisjon 1 er lik forventet poengsum gitt av førstesensor Bårdsen i kommisjon 3.

Andresensor Nansen har lang erfaring. Vi ønsker å teste om førstesensor Bårdsen og andresensor Nansen er enige om poenggivningen når de retter sammen i kommisjon 1.

- Test om forventet poengsum gitt av førstesensor og andresensor i kommisjon 1 er lik hverandre.

T-tester bygger på en antagelse om at variablene er normalfordelte. I store utvalg kan vi ved å støtte oss på sentralgrenseteoremet bruke T-test selv om variablene ikke er normalfordelte.

- d) Anta at *Poeng1* ikke er normalfordelt og ta stilling til om følgende utsagn er korrekte:
- (i) Med økende utvalgsstørrelse vil variabelen *Poeng1* nærme seg normalfordelingen.
  - (ii) Med økende utvalgsstørrelse vil gjennomsnittet til variabelen *Poeng1* nærme seg normalfordelingen.

Ett, begge eller ingen av utsagnene kan være riktige. Gi en kort begrunnelse for svaret.

Under er gitt to enveis variansanalyser av poenggivningen til første- og andresensorene. Variabelen *Sensor1* angir hvem som er førstesensor og variabelen *Sensor2* angir hvem som er andresensor.

**Analysis of Variance, response = Poeng1, treatment = Sensor1:**

	Sum of squares	df	Mean square
Treatment	0.0198045	1	0.0198045
Residual	99985.9	345	289.814
Total	99985.9	346	288.977

$F(1, 345) = 0.0198045 / 289.814 = 6.83352e-005$  [p-value 0.9934]

Level	n	mean	std. dev
Bårdsen	171	67.2459	16.398
Eriksen	176	67.2308	17.611

Grand mean = 67.2382

**Analysis of Variance, response = Poeng2, treatment = Sensor2:**

	Sum of squares	df	Mean square
Treatment	2443.57	2	1221.79
Residual	103930	344	302.123
Total	106374	346	307.439

$F(2, 344) = 1221.79 / 302.123 = 4.044$  [p-value 0.0184]

Level	n	mean	std. dev
Nansen	118	71.8983	19.345
Trondsen	114	65.6754	17.532
Kristensen	115	67.2348	14.930

Grand mean = 68.3084

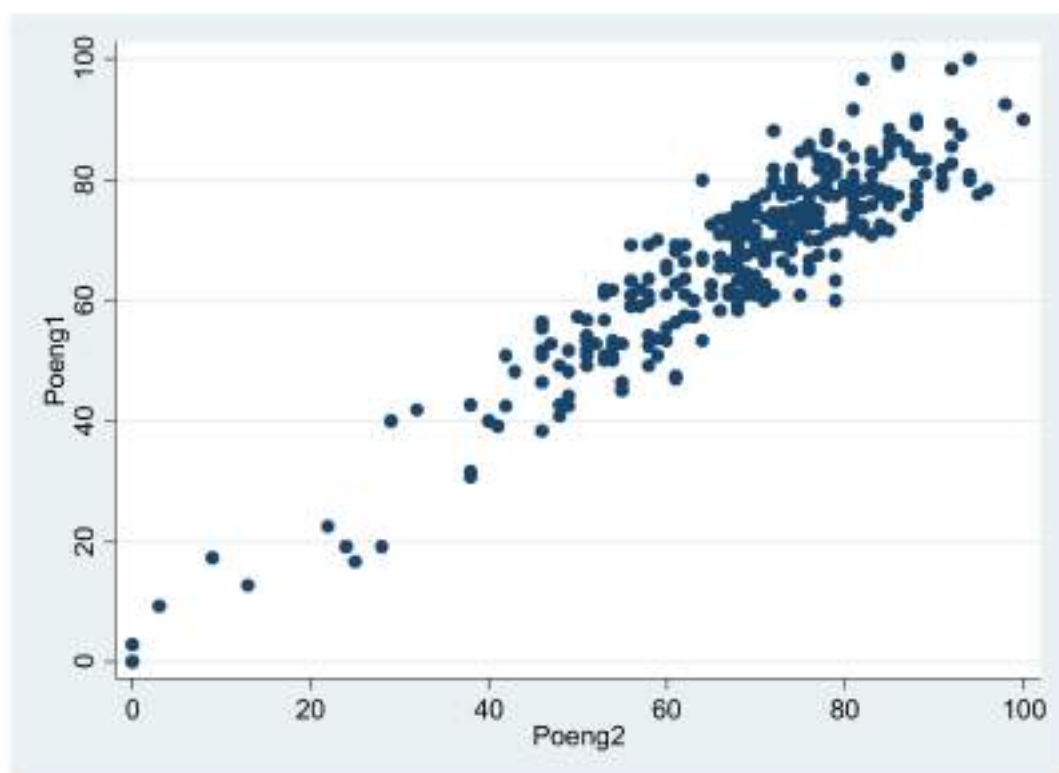
- e) (i) Hva testes i disse to variansanalysene? (ii) Oppgi to viktige forutsetninger for testen.  
(iii) Hva forteller testene oss?

Korrelasjonen mellom poengene til første- og andresensor (*Poeng1* og *Poeng2*) for alle 347 besvarelsene er **0.9306**.

- f) Test om denne korrelasjonskoeffisienten er signifikant forskjellig fra null. Kommenter kort svaret.
- g) Anta at den sanne korrelasjonskoeffisienten mellom *Poeng1* og *Poeng2* er 1. Hvilke av de to utsagnene under er da riktige?
  - (i) Sensorene er fullkomment enige om rangeringen av kandidatene.
  - (ii) Sensorene er fullkomment enige om poengene til kandidatene.

Ett, begge eller ingen av utsagnene kan være riktige. Gi en kort begrunnelse for svaret.

Under er det gjengitt et spredningsdiagram som viser første- og andresensors poeng på hver besvarelse. Førstesensors poeng (*Poeng1*) kan avleses på Y-aksen mens andresensors poeng (*Poeng2*) kan avleses på X-aksen.



Anta at du bruker minste kvadraters regresjon til å estimere regresjonslinjen  $Poeng1 = \alpha + \beta \cdot Poeng2$ .

- h) Regn ut forklaringsgraden til denne regresjonen.
- i) Regn ut stigningstallet til *Poeng2*.

Anta at det finnes en «sann» poengsum,  $P$ , som er det korrekte uttrykket for kandidatenes kunnskapsnivå. Anta videre at sensorene avviker tilfeldig fra denne sanne poengsummen på grunn av feilrettinger og ulikt skjønn som følger:

$$Poeng1 = P + \varepsilon_1$$

$$Poeng2 = P + \varepsilon_2$$

Variablene  $\varepsilon_1$  og  $\varepsilon_2$  representerer her målefeil. Vi antar at de to målefeilvariablene har samme varians, forventning null og at de er ukorrelerte både med hverandre og med den «sanne» poengsummen,  $P$ .

Korrelasjonskoeffisienten mellom *Poeng1* og *Poeng2* kalles i denne sammenhengen reliabilitetskoeffisienten og kan tolkes som den andelen av variansen til *Poeng1* og *Poeng2* som skyldes sann variasjon i  $P$ . I vårt tilfelle har vi da at 93 % av variansen til sensorenes poenggivning skyldes variansen i kandidatenes kunnskapsnivå, mens 7 % av variansen skyldes tilfeldigheter i sensorenes rettinger og skjønnsetning. For å bevise dette resultatet må man vise at følgende setning er korrekt, gitt modellen og forutsetningene ovenfor:

$$\text{Cov}(Poeng1, Poeng2) = \text{Var}(P). \quad (*)$$

- j) Ta utgangspunkt i definisjonen av varians og kovarians og vis at setningen (\*) gjelder.

Etter å ha satt poeng på besvarelsene hver for seg kommer første- og andresensor sammen og blir enige om en felles poengsum. I datasettet er denne variabelen kalt «Snitt» siden det normale er at sensorene bruker gjennomsnittet av sine respektive poengsummer. Dette vil redusere betydningen av tilfeldige avvik i bedømmelsene.

Under finner du resultatet av en regresjon med «Snitt» som avhengig variabel og dummier for første og andresensor som forklaringsvariabler. Dummyvariabelen *DSensor1\_B* er én dersom Bårdsen var førstesensor og null ellers. Dummyvariabelen *DSensor2\_T* er én dersom Trondsen var andresensor og null ellers. Dummyvariabelen *DSensor2\_K* er én dersom Kristensen var andresensor og null ellers.

Model 1: OLS, using observations 1-347  
Dependent variable: Snitt

	Coefficient	Std. Error	t-ratio	p-value	
Const	68.6787	1.80358	38.0790	<0.00001	***
DSensor1_B	1.58399	1.8174	0.8716	0.38405	
DSensor2_T	-4.23241	2.22273	-1.9041	0.05773	*
DSensor2_K	-0.960607	2.2176	-0.4332	0.66516	
Mean dependent var	67.77330	S.D. dependent var		16.96438	
Sum squared resid	98237.68	S.E. of regression		16.92357	
R-squared	0.013435	Adjusted R-squared		0.004806	
F(3, 343)	1.556951	P-value(F)		0.199624	

- k) Forklar hvorfor det er tilstrekkelig med tre dummyvariabler når det er fem sensorer involvert i rettingen.
- l) Tolk kort regresjonsutskriften.

Sensorene blir sammen enige om karaktergrensene. Tabellen under gir den endelige karakterfordelingen for de 333 studentene som bestod eksamen. Tallkarakter 5 er A, 4 er B, osv.

**Frequency distribution for Tallkarakter, obs 1-333**

Subsampled data (restriction: Tallkarakter>0)

	frequency	rel.	cum.	
1 (E)	4	1.20%	1.20%	
2 (D)	41	12.31%	13.51%	****
3 (C)	91	27.33%	40.84%	*****
4 (B)	127	38.14%	78.98%	*****
5 (A)	70	21.02%	100.00%	*****

I en Wikipedia-artikkel om karaktersystemet i Norge kan vi lese at karakterskalaen ovenfor er «basert på det omdiskuterte ECTS-systemet som baserer seg på at de 10 % beste (over tid) besvarelsene skal gis karakter A, de 25 % nest beste får B, de midterste 30 % får C, de 25 % nest dårligste får D og de 10 % dårligste får E. Prosentene er blant de som ikke stryker og får karakter F.»

- m) Test om karakterfordelingen i INTO10 høsten 2007 skiller seg signifikant fra den fordelingen som er oppgitt i Wikipedia-artikkelen.

Det var 4 % av kandidatene som strøk våren 2007. Anta at dette er den sanne stryksannsynligheten for en tilfeldig kandidat som skal opp til eksamen våren 2015 og at det er 400 kandidater som går opp til eksamen.

- n) Beregn et 95 % konfidensintervall for strykprosenten våren 2015.

Variabelen *Snitt* har i 2007-dataene et gjennomsnitt på 68 og et standardavvik på 17. Anta at disse verdiene representerer sann forventning og sant standardavvik for fordelingen. Strykgrensen var 30 poeng.

- o) Hvilken strykprosent vil vi forvente å observere i 2007-dataene dersom *Snitt* er eksakt normalfordelt?