



Høgskolen i Telemark
Fakultet for allmennvitenskapelige fag

EKSAMEN

6019
Statistikk II

9.12.2014

Tid: 9-13 (4 timer)

Målform: Bokmål

Sidetal: 4 (inkludert denne)

Hjelpemiddel: Alle trykte og skrevne samt kalkulator.

Merknader: Ingen

Vedlegg: Ingen.

Sensuren finner du på StudentWeb.

Oppgave 1

Det blir ofte hevdet at jenter i større grad flytter fra landsbygda enn gutter. For å belyse dette spørsmålet er det foretatt en undersøkelse i bygdekommuner i Sør-Norge som omfattet totalt 1700 ungdommer, 912 jenter og 788 gutter. Oversikt over hvor mange som flyttet fra landsbygda og hvor mange som ble boende for begge kjønn er gitt i tabellen nedenfor.

	Flyttet	Ble boende	Sum
Jenter	343	569	912
Gutter	233	555	788
Sum	576	1124	1700

- a) Tyder disse tallene på at det er forskjell mellom kjønnene når det gjelder spørsmålet om en velger å bli boende på bygda eller å flytte ut? Formuler dette spørsmålet som en hypotesetest. Gjennomfør testingen og angi konklusjonen når signifikansnivået velges lik 1 %.

Oppgave 2

Fjellmegleren i Telemark har trukket ut 12 hytter i Rauland som er lagt ut til salg på *finn.no*. Vi antar at prisen på en hytte X i tusen kroner er en kontinuerlig fordelt stokastisk variabel. Sannsynlighetstettheten, f_X , er symmetrisk, og medianen i fordelingen er m_1 . Vi antar dessuten at prisene på forskjellige hytter er uavhengige variabler. Resultatene er gitt til slutt i oppgaven (neste side).

- a) Fjellmegleren vil teste om medianen i fordelingen for hytteprisene i Rauland er under 3.0 millioner kroner (dvs. 3000 tusen kroner). Formuler dette spørsmålet som en hypotesetest om m_1 . Test hypotesen ved å bruke Wilcoxon's ett-utvalgstest. Angi konklusjonen når signifikansnivået velges lik 5 %.

Fjellmegleren i Telemark vil dessuten teste om prisnivået på hytter er lavere i Rauland enn i Hafjellsområdet. Firmaet har derfor også trukket ut 10 hytter fra Hafjellsområdet som ligger til salg på *finn.no*. Vi antar at sannsynlighetstettheten for hyttepriser i Hafjell har samme form som f_X , og at prisene på forskjellige hytter er uavhengige variabler.

- b) Tyder disse hytteprisene som er innhentet fra de to populære vinterdestinasjonene, at prisnivået på hytter er lavere i Rauland enn i Hafjellsområdet? Formuler dette spørsmålet som en hypotesetest. Gjennomfør testingen og angi konklusjonen når signifikansnivået velges lik 1 %.

Hvilken test kunne du ha brukt om du skulle undersøke om det er forskjeller i prisnivået på hytter i Rauland, Hafjell og Trysil?

Resultater

Priser på de til sammen 22 hyttene i tusen kroner (sortert etter stigende pris) for hver av de to hytteområdene:

Rauland: 1790 2190 2250 2380 2390 2390 2475 2570 2990 3240 4200 5500

Hafjell: 2430 2900 2950 3750 3925 3975 4690 5495 7290 16000

Oppgave 3

Vi skal studere hvordan ansiennitet, utdanning, lederansvar og kjønn virker inn på medarbeidernes lønn i en stor industribedrift. 22 ansatte i bedriften er trukket ut, og for hver ansatt er det i tillegg til årslønnen Y registrert følgende variabler:

- X_1 = " Antall år i bedriften (= ansiennitet)"
- X_2 = " Antall år med høyere utdanning"
- X_3 = " Antall personer den ansatte er leder for (=lederansvar)"
- X_4 = " Kjønn (0 betyr kvinne og 1 mann)"

Vi skal bruke en multippel regresjonsmodell til å studere sammenhengen mellom årslønnen Y (i kroner) og de 4 forklaringsvariablene X_1 , X_2 , X_3 og X_4 :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + U$$

Observasjonene og regresjonsanalysen går fram av Minitab-utskriften til slutt i oppgaven (neste side). Studer denne før du svarer på spørsmålene nedenfor.

- a) Gjør kort greie for antakelsene i regresjonsmodellen. Skriv opp den estimerte regresjonsmodellen.

Prediker årslønnen til en kvinne med 5 års ansiennitet, 4 års høyere utdanning som er leder for 7 ansatte. Forklar kort hva den estimerte verdien av β_0 uttrykker.

Forklar dessuten kort hva R^2 (R -sq i Minitab-utskriften) uttrykker.

- b) Hva er estimert forventet forskjell i årslønn mellom en mann og en kvinne som har lik ansiennitet, lik utdanning og likt lederansvar. Angi et 95 % konfidensintervall for denne forskjellen.
- c) Hvilke av de 4 forklaringsvariablene har signifikant innvirkning på årslønnen? Formuler selv aktuelle hypoteser og angi konklusjoner på testene. Bruk signifikansnivå 5 %.

Blant de 22 ansatte som er trukket ut til å være med i denne analysen, er det 12 menn og 10 kvinner. Hvis man her gjennomfører en to-utvalgs t -test for sammenligning av to grupper i målemodellen for å undersøke om det er forskjell i forventet årslønn for menn og kvinner, blir t -verdien $T_0 = 2.28$. (Dette kan tas for gitt.)

Gjør kort greie for at konklusjonen på denne t -testen er at vi kan påstå at det er forskjell i forventet årslønn mellom menn og kvinner på 5 % nivå. Kan du gi en forklaring på hvorfor denne t -testen og testen av regresjonskoeffisienten som uttrykker forskjell mellom kjønnene, gir ulike konklusjoner?

Utskrift fra Minitab

Data Display

Row	Y	X1	X2	X3	X4
1	769000	18	6	5	1
2	737000	15	6	4	0
3	675000	5	1	0	1
4	627000	6	0	1	0
5	778000	22	5	7	1
6	685000	3	2	6	1
7	747000	21	3	9	1
8	778000	18	3	5	0
9	670000	0	3	4	0
10	658000	0	5	0	1
11	810000	6	2	2	1
12	728000	3	4	3	0
13	680000	2	2	5	0
14	732000	9	1	2	0
15	858000	12	5	6	1
16	745000	6	2	2	1
17	734000	9	0	5	0
18	645000	0	0	2	0
19	835000	15	4	4	1
20	628000	0	0	2	0
21	728000	19	2	6	1
22	744000	4	5	8	1

MTB > Regress 'Y' 4 'X1'-'X4';

Regression Analysis: Y versus X1; X2; X3; X4

The regression equation is
 $Y = 653038 + 3728 X1 + 8160 X2 + 1287 X3 + 24477 X4$

Predictor	Coef	SE Coef	T	P
Constant	653038	22354	29,21	0,000
X1	3728	1805	2,07	0,054
X2	8160	6279	1,30	0,211
X3	1287	5364	0,24	0,813
X4	24477	23297	1,05	0,308

S = 48764,7 R-Sq = 52,5% R-Sq(adj) = 41,3%