

## Eksamen i STAT100- statistikk grunnkurs

Bokmål

Mandag 6. desember 2004

C2, alle trykte og skrevne hjelpemidler, samt lommekalkulator.

Ved vurdering teller alle delpunkter (1a, 1b etc) likt.

Oppgaveteksten er på 4 sider

Kontakt under eksamen: Trygve Almøy, tlf: 8867

### Oppgave 1

To takstmenn (Olsen og Hansen) takserer boligeiendommer. En har mistanke om at Olsen systematisk takserer høyere enn Hansen. En trakk tilfeldig ut 6 eiendommer, og begge takstmenn takserte disse uten å vite om hverandre.

Her er resultatene fra Olsen (i millioner kroner og rundet av til nærmeste 10 000)

Eiendom nr	1	2	3	4	5	6
Takst	1,44	1,59	1,02	1,67	0,98	2,54

Her er resultat fra Hansen (i millioner kroner og rundet av til nærmeste 10 000)

Eiendom nr	1	2	3	4	5	6
Takst	1,35	1,57	1,09	1,54	0,88	2,45

a) Gir dette grunnlag for å påstå at Olsen generelt gir høyere takst enn Hansen? Formuler problemstillingen som en hypotesetest. Utfør testingen og trekk konklusjon.

b) I en annen undersøkelse ga Olsen høyere takster enn Hansen på 30 av 40 eiendommer. Bruk denne undersøkelsen som grunnlag for å teste om Olsen generelt gir høyere takst enn Hansen.

### Oppgave 2

Det ble foretatt en verdensomspennende undersøkelse om helse og levealder. I utgangspunktet var alle land med, men vi har plukket ut 13 land.

Dette gav følgende:

Land	Levealder	Helseutgift
Cameroon	48,1	42
Colombia	71,8	356
DominicanRep	68,0	353
Gambia	57,1	78
Indonesia	66,4	77
Lesotho	35,7	101
Malaysia	72,0	345
Mongolia	62,9	122
Niger	42,6	22
Samoa	68,2	199
SierraLeone	34,0	26
Turkey	70,0	294
Zambia	39,7	52

La responsvariabel (Y) være levealder og la forklaringsvariabel (x) være helseutgift, målt i innsats pr person (i 2001).

Vi antar følgende modell.

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i,$$

der  $\varepsilon_i$ -ene er uavhengige og normalfordelt med forventning 0 og standardavvik  $\sigma$ .

Vi kjørte modellen i Minitab og fikk følgende resultat:

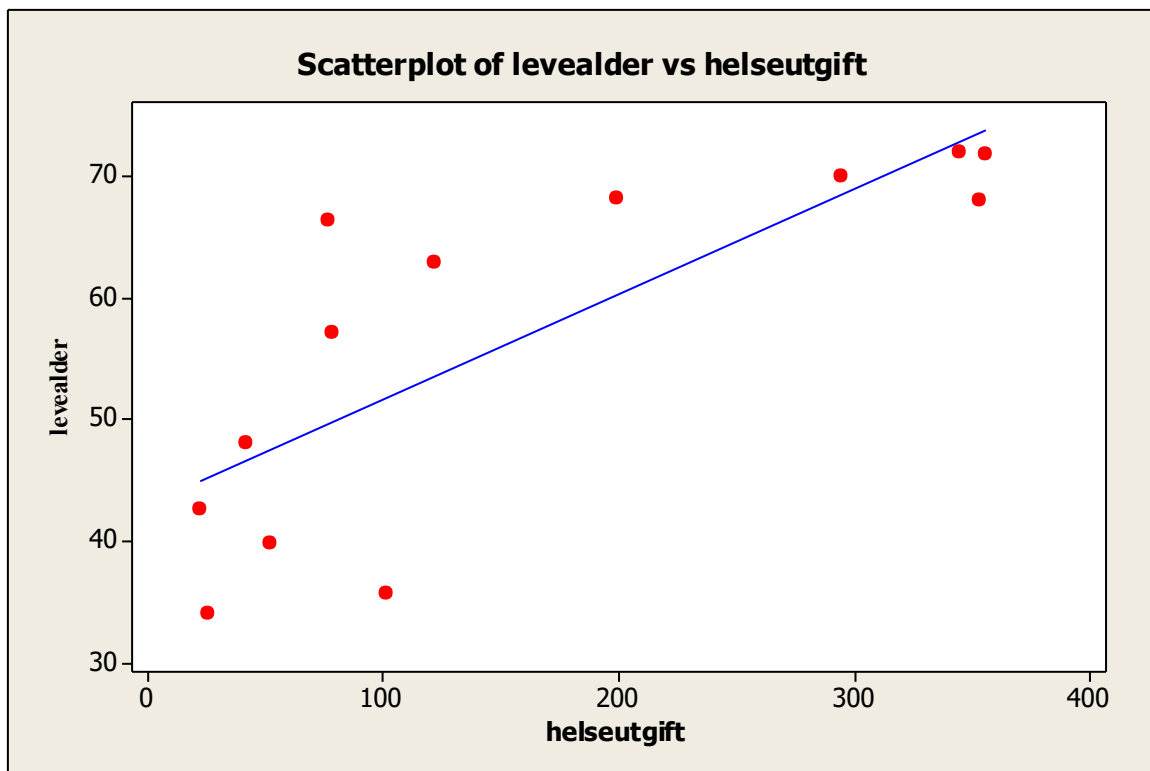
Regression Analysis: levealder versus helseutgift					
The regression equation is					
levealder = * + *					
Predictor	Coef	SE Coef	T	P	
Constant	42,901	4,179	10,27	0,000	
helseutgift	0,086	0,020	*	0,001	
S = * R-Sq = * R-Sq(adj) = 58,3%					
Analysis of Variance					
Source	DF	SS	MS	F	P
Regression	1	1574,6	1574,6	17,76	0,001
Residual Error	11	975,1	88,6		
Total	12	2549,7			

a) Gi en tolkning av alle parametrene i modellen, og finn de estimerte verdiene. En del av verdiene i utskriften er erstattet med en \*. Fyll ut de riktige verdiene.

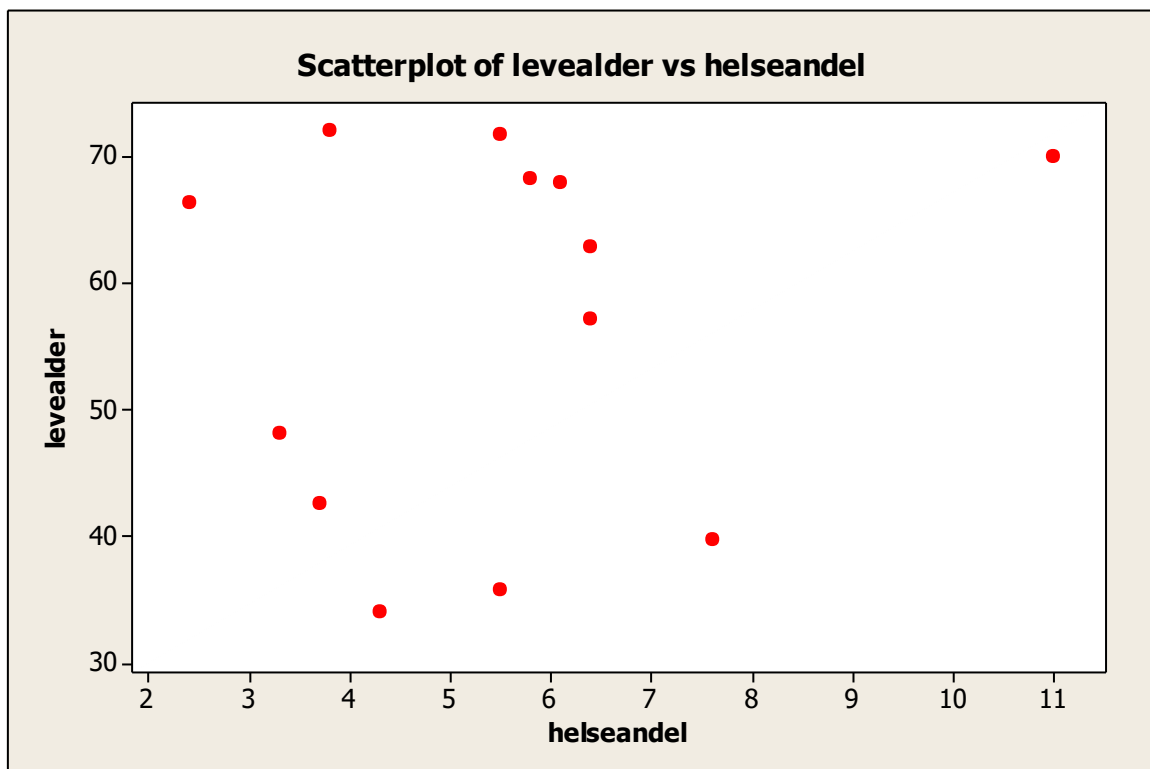
b) Lag et 99 prosent konfidensintervall for  $\beta_1$ , og forklar hva dette intervallet sier.

c) Peru brukte 300 dollar pr innbygger på helse i 2001. Prediker levealder i Peru. Peru har et 95 % prediksjonsintervall for levealder på **(46.42, 91.28)**. Forklar hva dette betyr. Italia brukte 2200 dollar pr innbygger i 2001. Hvorfor kan ikke denne modellen brukes til å predikere levealder i Italia?

d) På Figur 1 ser du et scatterplott med inntegnet estimert regresjonslinje. Ser det ut til at modellen passer til data? Hvordan tror du et residualplot vil se ut? (Bare skisser det). Vi prøvde også å bruke "Andel av brutto nasjonalprodukt brukt på helse" som forklaringsvariabel i stedet for "Helseutgift pr person". Et plott av dette ser du i Figur 2. Tror du dette er en brukbar forklaringsvariabel? Hva tror du om størrelsen på  $R^2$  her?



Figur 1. Levealder mot helseutgifter med inntegnet regresjonslinje



Figur 2: Levealder mot andel av brutto nasjonalprodukt brukt på helse

### Oppgave 3

La en ku's årlig produksjon av mjølkefett være normalfordelt med forventning ( $\mu$ ) 170 kg og standardavvik ( $\sigma$ ) 20 kg.

a) Hva er sannsynligheten for at en tilfeldig valgt ku produserer mer enn 190 kg. mjølkefett? En definerer en ku som høytstående dersom den er blant de 10% med høyest mjølkefettproduksjon. Hvor mye må en ku minst produsere for å regnes blant de høytstående?

I et avlsforsøk ble fire okser som alle hadde svært mange døtre undersøkt. For hver av de fire oksene trakk vi ut fire døtre og registrerte døtrenes årlige produksjon av mjølkefett.

Okse nr	1	2	3	4
	179	159	118	207
	143	151	220	245
	112	169	128	189
	158	132	154	201

b) Okse 4 er relativt ung, og en vurderer om denne skiller seg positivt ut. Kan du med rimelig sikkerhet påstå at det er positiv forskjell på denne og de tre andre? (Bruk utskriften under så mye du kan).

#### One-way ANOVA:

Analysis of Variance

Source	DF	SS	MS	F	P
Factor	3	10398	3466	3,71	0,042
Error	12	11198	933		
Total	15	21596			

Okse	N	Mean
1	4	148,00
2	4	152,75
3	4	155,00
4	4	210,50

Pooled StDev = 30,55