

# UNIVERSITETET I OSLO

## Det matematisk-naturvitenskapelige fakultet

Eksamen i: STK 1000 — Innføring i anvendt statistikk.

Eksamensdag: Torsdag 1. juni 2006.

Tid for eksamen: 09.00 – 12.00.

Oppgavesettet er på 5 sider.

Vedlegg: Tabell over normalfordeling, tabell over binomisk fordeling.

Tillatte hjelpemidler: Lærebok: Moore & McCabe "Introduction to the practice of statistics", 3., 4. el. 5. utgave. Ordliste for bruk i STK 1000, kalkulator.

Kontroller at oppgavesettet er komplett før du begynner å besvare spørsmålene.

### Oppgave 1.

Det fødes omtrent like mange gutter som jenter, men i store befolkninger er det likevel ofte mulig å demonstrere at det er et avvik fra 50% av hvert kjønn som ikke kan tilskrives tilfeldigheter. I denne oppgaven beskrives en modell og en metode for påvise dette.

- a) La  $X$  = antall gutter i  $n = 10$  fødsler. Anta at  $p$  er sannsynligheten for gutt i en fødsel. Er det rimelig å anta at  $X$  er binomisk fordelt? Begrunn svaret.
- b) Anta at  $X$  er binomisk fordelt med  $n = 10$  og  $p = 0.5$  ( $X$  er  $B(10, 0.5)$ ). Finn (fra tabell)
  - i) Sannsynligheten for at  $X = 5$
  - ii) Sannsynligheten for at  $X \leq 4$
- c) Anta vi observerer kjønnet for  $n = 10,000$  barn og finner  $X = 5143$  gutter. Test på denne bakgrunn  $H_0 : p = 0.5$  mot  $H_a : p \neq 0.5$ . Beregn testobservator og  $p$ -verdi for testen. Konkluder.

(Fortsettes side 2.)

## Oppgave 2.

Denne oppgaven dreier seg om et datasett over ville bjørner, som er blitt bedøvet og deretter veid og målt på ulike måter. Et formål med studien er å kunne si noe om bjørners vekt basert på data om kroppsstørrelse og kjønn. Dette kan være nyttig ved framtidige observasjoner av bjørner som kan måles, men hvor en vekt ikke er tilgjengelig. I denne oppgaven skal vi bare benytte bjørnenes vekt (**Weight**, målt i pund), kjønn (**Sex**, kodet ved 1 = hann og 2 = hunn) samt deres brystomfang (**Chest.G**, målt i tommer).

- a) Under er det satt opp en Minitab-utskrift fra en hypotesetest for om forventet vekt er den samme blant hunn-bjørner som blant hann-bjørner. Still opp nullhypotesen og den alternative hypotesen for problemet. Formuler en konklusjon for testen.

Two-Sample T-Test and CI: Weight; Sex

Two-sample T for Weight

Sex	N	Mean	StDev	SE Mean
1	99	214	120	12
2	44	143.0	64.5	9.7

Difference = mu (1) - mu (2)

Estimate for difference: 70.9444

95% CI for difference: (40.3627; 101.5262)

T-Test of difference = 0 (vs not =): T-Value = 4.59 P-Value = 0.000 DF = 135

- b) To-utvalgstester som den i punkt a) kan gjøres både med antagelse om at standardavvikene i gruppene er like og med antagelse om at standardavvikene er ulike.  
Hvilken test ble benyttet i punkt a)?  
Synes du riktig test ble benyttet? Begrunn svaret.
- c) Utskriften angir også et konfidensintervall for forskjell i forventet vekt mellom hunn-bjørner og hann-bjørner.  
Vis hvordan konfidensintervallet er regnet ut.  
Gi en fortolkning av dette konfidensintervallet.

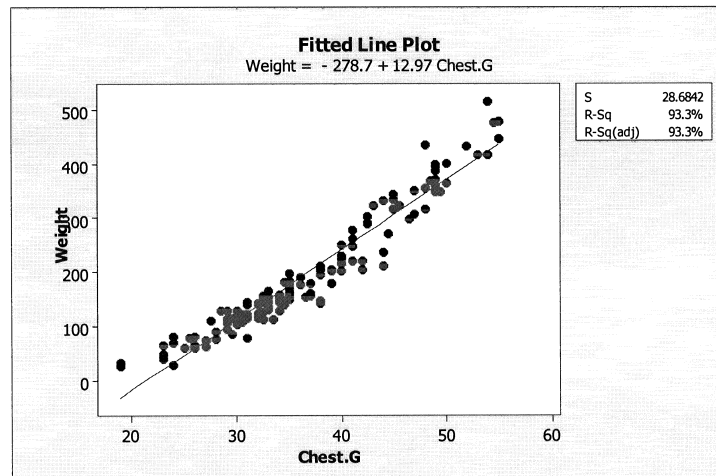
I de to neste punktene skal vi studere hvordan bjørners vekt avhenger av deres brystomfang, men uten å ta hensyn til kjønnsforskjeller. I oppgavens siste punkt trekkes også kjønn inn i analysen.

(Fortsettes side 3.)

- d) Under er det satt opp resultater fra en enkel lineær regresjon av  $x =$  bjørnens omkrets rundt brystkassa mot  $y =$  bjørnens vekt.  $R^2$  er angitt til 93.3%.

Forklar og diskuter begrepet  $R^2$ .

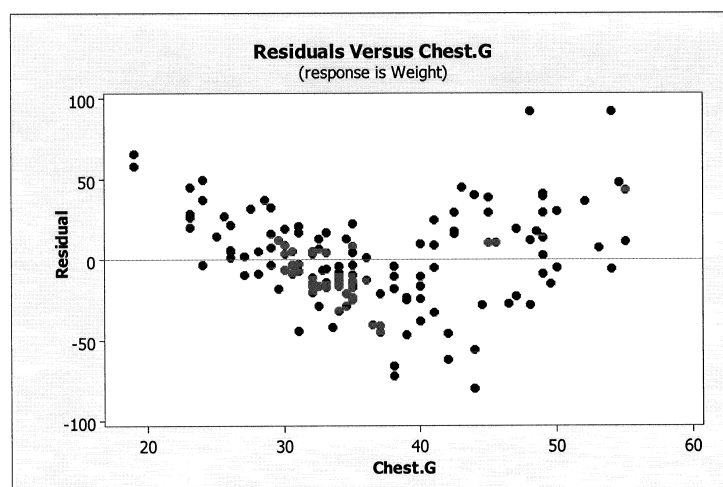
Kommenter om  $x$ -variablens evne til å predikere vekten for en ny bjørn.



- e) Under er det satt opp et plott av residualene fra regresjonsanalysen i punkt d) mot  $x$ -variablen.

Forklar hva residualer er og angi hvordan disse kan benyttes til å avsløre modellavvik.

Forklar spesielt hvorfor det ser ut til å være avvik fra linearitet i dette tilfelle.



- f) Under er det satt opp resultater fra en multippel regresjon med forklaringsvariable

$x_1 = x$  = lengden av bjørnens omkrets rundt brystkassa,

$x_2 = x^2$  = lengden av bjørnens omkrets rundt brystkassa kvadrert

I Minitab er kvadratleddet betegnet med **Chest2**. Resultatene viser både den tilpassede sammenhengen mellom  $x$  og  $y$  og et residualplott mot  $x$ -variablen.

Skriv opp modellen for denne regresjonsanalysen og identifiser estimatene for parametrene som inngår.

Indikerer residualplottet fortsatt modellavvik? Begrunn svaret.

Regression Analysis: Weight versus Chest.G; Chest2

The regression equation is

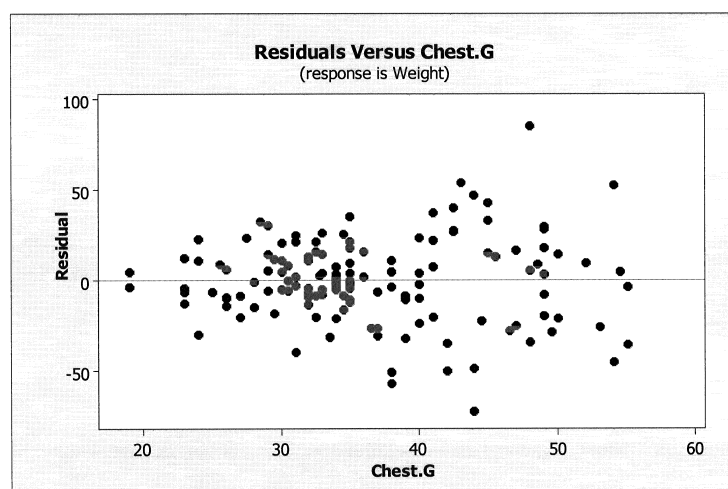
Weight = 12.7 - 3.17 Chest.G + 0.212 Chest2

Predictor	Coef	SE Coef	T	P
Constant	12.65	35.19	0.36	0.720
Chest.G	-3.166	1.901	-1.67	0.098
Chest2	0.21247	0.02484	8.56	0.000

S = 23.3276    R-Sq = 95.6%    R-Sq(adj) = 95.5%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	1659068	829534	1524.38	0.000
Residual Error	140	76185	544		
Total	142	1735253			



- g) I en siste regresjon inkluderes i tillegg forklaringsvariablen  $x_3 =$  kjønn for bjørnen (kodet ved 1 = hann og 2 = hunn). Resultatet av regresjonen er gitt i tabellen på neste side. Test om forventet vekt er

(Fortsettes side 5.)

den samme blant hunn-bjørner som blant hann-bjørner under denne regresjonsmodellen.

Sammenhold resultatet med testen i punkt a) og gi en forklaring på forskjellen. Du kan benytte at korrelasjonskoeffisienten mellom variablene Chest.G og Sex er lik  $-0.26$ .

Regression Analysis: Weight versus Chest.G; Chest2; Sex

The regression equation is

Weight = 21.1 - 2.93 Chest.G + 0.208 Chest2 - 8.07 Sex

Predictor	Coef	SE Coef	T	P
Constant	21.09	35.19	0.60	0.550
Chest.G	-2.930	1.889	-1.55	0.123
Chest2	0.20780	0.02475	8.40	0.000
Sex	-8.074	4.362	-1.85	0.066

S = 23.1282 R-Sq = 95.7% R-Sq(adj) = 95.6%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	1660900	553633	1035.00	0.000
Residual Error	139	74353	535		
Total	142	1735253			

SLUTT