

# UNIVERSITETET I OSLO

## Det matematisk-naturvitenskapelige fakultet

Eksamen i                   STK1000 — Innføring i anvendt statistikk  
Eksamensdag:            Torsdag 2. desember 2010.  
Tid for eksamen:        09.00 – 13.00.  
Oppgavesettet er på 6 sider.  
Vedlegg:                  Ingen  
Tillatte hjelpemidler:  Lærebok: Moore & McCabe: Introduction to  
                                  the practice of statistics,  
                                  ordliste for bruk i STK1000, godkjent kalkulator

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

### Oppgave 1

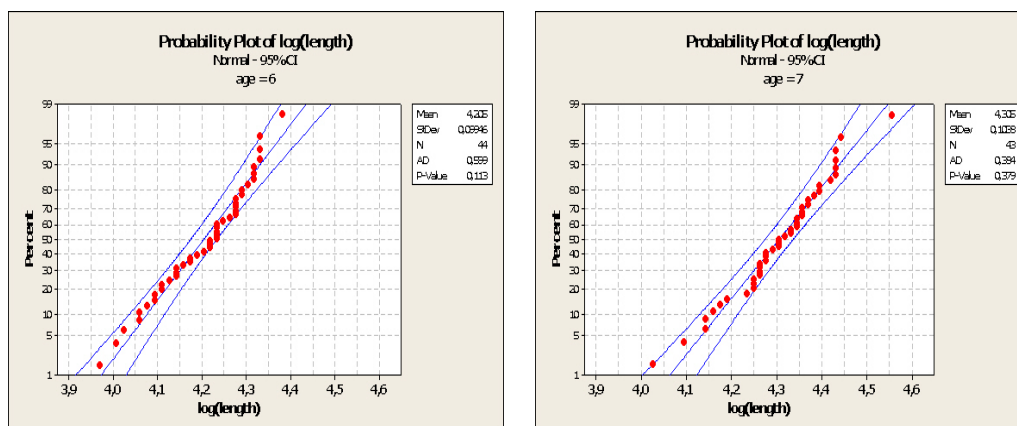
Bestandsestimering og setting av fangstkvoter er nødvendig for å opprettholde ulike fiskepopulasjoner på bærekraftig nivå. Bestandsestimering baseres på innsamling av data og (tildels avansert) statistisk analyse. Da kvoter gjerne inneholder restriksjoner med hensyn på lengde av fisk, mens utvikling av en populasjon er svært avhengig av alderssammensetningen innen populasjonen, er en viktig komponent av en slik analyse sammenhengen mellom alder og lengde av fisk. Vi vil i denne og neste oppgave se på slike sammenhenger. Generelt vil fisk vokse med årene men vekstkurven vil ofte flate ut ved høyere aldre.

Data er samlet inn ved at 133 fisk (torsk i dette tilfellet) er tilfeldig utvalgt fra den totale fangst fra en båt og for hver av fiskene er lengde og alder målt (hvordan alder blir målt vil være tema for oppgave 3).

I denne oppgaven vil vi konsentrere oss om fisk i alder 6 eller 7 år som omfatter tilsammen 87 av de 133 innsamlede fisk. Tabellen nedenfor viser oppsummerende statistikk for lengde for de to alderskategoriene. Her er en log (naturlig logaritme) transformasjon gjort på lengde dataene som opprinnelig var gitt i cm. Under de oppsummerende statistikkene er også kvantilplott av (log) lengde gitt for de to alderskategoriene.

Variable	age	N	Mean	SE Mean	StDev	Min	Q1	Median	Q3	Max
log(length)	6	44	4,205	0,015	0,100	3,970	4,131	4,227	4,277	4,382
	7	43	4,305	0,016	0,103	4,025	4,245	4,304	4,369	4,554

(Fortsettes på side 2.)



- (a) Beskriv de modell antagelser som ligger til grunn for en to-utvalgs t-test og diskutér basert på oppsummerende statistikk og kvantilplott om disse antagelser er rimelige i denne situasjonen.

Hva er populasjonene i dette tilfellet?

Nedenfor er en Minitab utskrift fra en to-utvalgs t-test basert på dette datasettet.

Two-sample T for log(length)

age	N	Mean	StDev	SE Mean
6	44	4,2047	0,0995	0,015
7	43	4,305	0,104	0,016

Difference = mu (6) - mu (7)

Estimate for difference: -0,1005

95% CI for difference: (-0,1438; -0,0571)

T-Test of difference = 0 (vs <):

T-Value = -4,61 P-Value = 0,000 DF = 84

- (b) Sett opp en null og en alternativ hypotese for problemstillingen og som svarer til utskriften ovenfor.

Diskutér hvorfor et ensidig alternativ er fornuftig i dette tilfellet.

- (c) Skriv opp formelen for den relevante test-observatoren og forklar hvordan P-verdien er beregnet (du behøver ikke her å gjøre de faktiske beregninger, kun forklare prosedyren).

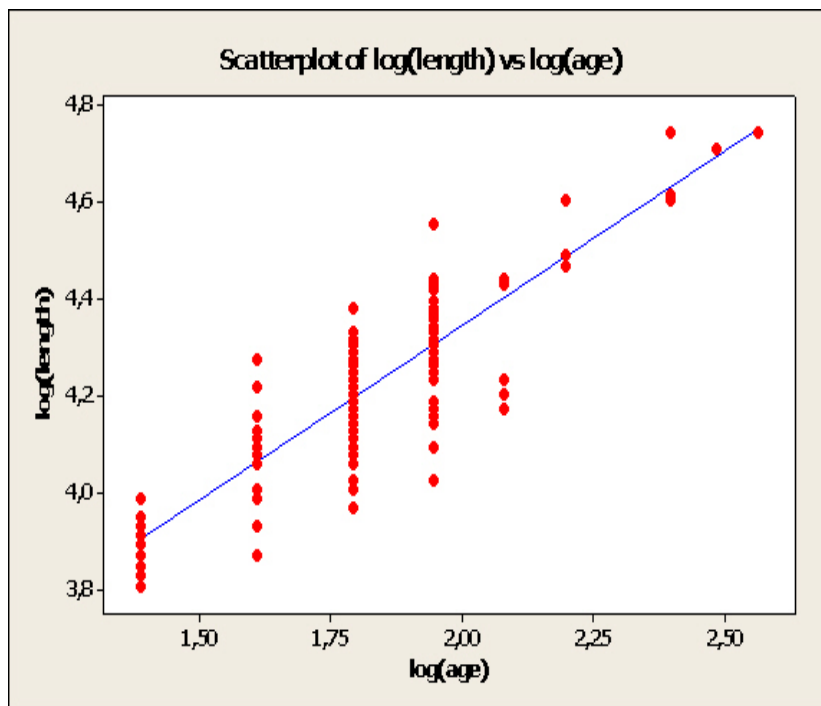
Bruk utskriften til å utføre testen samt å formulere en konklusjon på testen.

## Oppgave 2

Vi vil fortsette i denne oppgaven med å se på sammenheng mellom alder og lengde av fisk. Denne oppgaven kan imidlertid løses uavhengig av den

(Fortsettes på side 3.)

foregående oppgaven. Plottet nedenfor viser lengde mot alder, begge variable på log-skala.



Basert på dette plottet og tidligere studier, er en rimelig modell for sammenhengen mellom lengde og alder av fisk gitt ved

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

der  $y_i$  er lengde for fisk  $i$  (på log-skala) mens  $x_i$  er alder (i år) til fisk  $i$  (også gitt på log-skala). Nedenfor er også en utskrift fra Minitab basert på denne regresjonsmodellen for totalt  $n = 133$  fisk.

General Regression Analysis: log(length) versus log(age)

Regression Equation

$$\log(\text{length}) = 2,91226 + 0,716424 \log(\text{age})$$

Coefficients

Term	Coef	SE Coef	T	P
Constant	2,91226	0,0679953	42,8303	0,000
log(age)	0,71642	0,0370862	19,3178	0,000

Summary of Model

$$S = 0,0974918 \quad R\text{-Sq} = 74,16\%$$

(Fortsettes på side 4.)

## Predicted Values for New Observations

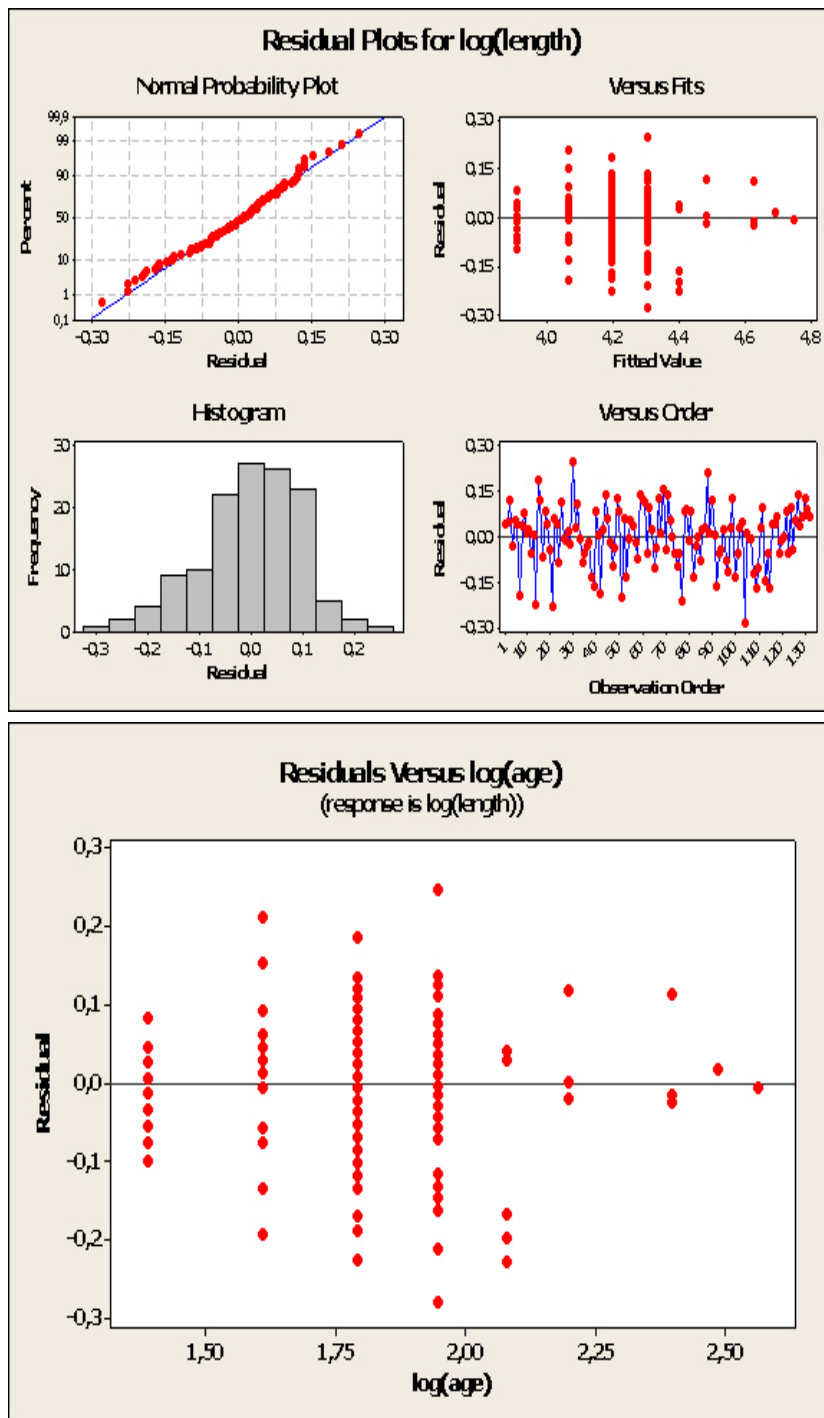
New Obs	Fit	SE Fit	95% CI	95% PI
1	4,48641	0,0163905	(4,45398; 4,51883)	(4,29082; 4,68199)

## Values of Predictors for New Observations

New Obs	log(age)
1	2,19723

- (a) Spesifiser hva parametrene er i modellen, hva de betyr og angi for hver parameter et estimat for disse basert på Minitab utskriften ovenfor.
- (b) Forklar hva R-Sq i utskriften er og hva den sier om modellen.  
Antyder den at vi har en god eller dårlig modell i dette tilfellet?
- (c) Lag et 95% konfidensintervall for  $\beta_1$ .  
Hvordan kan P-verdien relatert til  $\beta_1$  i Minitab utskriften brukes til å forklare at intervallet ikke dekker 0?
- (d) Anta vi er interessert i lengde på fisk for alder 9 år som svarer til at  $x = \log(9) = 2.19723$ . Den siste delen av Minitab utskriften omhandler dette.  
Forklar hva som er forskjellen på det oppgitte konfidensintervallet (CI) i forhold til det oppgitte prediksjonsintervallet (PI).  
Gi en forklaring på hvorfor prediksjonsintervallet er bredere enn konfidensintervallet.  
Gitt at modellen vår baserer seg på data på log-skala, angi konfidensintervall og prediksjonsintervall på opprinnelig skala.
- (e) Nedenfor er det vist ulike residualplott basert på regresjonsmodellen ovenfor.  
For hver figur, forklar hvilke antagelser som kan sjekkes og diskutér om antagelse er rimelige.

(Fortsettes på side 5.)



(Fortsettes på side 6.)

### Oppgave 3

Spesifisering av alder på fisk er ikke en enkel affære. Aldersspesifisering for fisk gjøres ganske tilsvarende aldersspesifisering for trær ved at fiskens *otolitter* (ørelignende strukturer) får en ny ring per år, slik at en ren optelling av ringer kan utføres. Slike ringer kan imidlertid i noen tilfeller være uklare, slik at aldersspesifisering kan være usikker.

Anta vi vet at sannsynlighetsfordelingen for den *sanne* alder for en tilfeldig fisk er gitt ved følgende fordeling:

Alder	$\leq 3$	4	5	6	$\geq 7$
Sannsynlighet	?	0.10	0.14	0.33	0.42

Her betyr  $\leq 3$  alle aldre mindre eller lik 3 og  $\geq 7$  betyr alle aldre større eller lik 7.

- (a) Hva er sannsynligheten for at en fisk har alder mindre eller lik 3 år?

Anta nå at hvis en fisk har sann alder  $a$ , så er sannsynligheten for at en *anslår* alderen til å være lik den sanne verdien  $a$  lik 0.95. Anta også at sannsynligheten for (feilaktig) å anslå verdien til  $a - 1$  er 0.05 når den sanne alder er  $a$  (dvs det er en viss sjanse for å ikke se en av ringene).

- (b) Anta en fisk *anslås* å ha alder 5 år. Finn sannsynligheten for at den *sanne* alder også er 5 år.

Hint: Argumentér først for at de eneste mulige sanne aldre er 5 eller 6 år.

SLUTT