



EKSAMENSOPPGAVER

Institutt: IKBM

Eksamen i: STAT100 STATISTIKK

Tid: Mandag 23.mai 2011 09.00-12.30 (3.5 timer)

Emneansvarlig: Solve Sæbø (Tlf: 90065281)

Tillatte hjelpemidler: C3: alle typer kalkulator, alle andre hjelpemidler

Opgaveteksten er på: 5 (inkludert F-tabell)
antall sider inkl. vedlegg

Alle deloppgaver teller likt.

Alle steder der du utfører statistiske tester, skal du skrive opp både nullhypotesen og den alternative hypotesen.

Bestandsestimering og setting av fangstkvoter er nødvendig for å opprettholde ulike fiskepopulasjoner på bærekraftig nivå. Bestandsestimering baseres på innsamling av data og statistisk analyse. Fiskekvoter inneholder restriksjoner med hensyn på lengde av fisk, og en viktig del av analysen er derfor å estimere hvordan lengden på fisk varierer med alder og fra sesong til sesong. Data på lengde og alder ble samlet inn på totalt 400 torsk gjennom fire ulike sesonger. Alderen på torskene ble bestemt ut fra antall årringer i *otolittene* i ørene, og i dataene varierte alderen mellom 2 og 13 år. Disse dataene skal vi bruke i hele eksamensoppgaven.

Nedenfor er lengde på fisk (cm) for aldrene 2, 3 og 4 år oppsummert med gjennomsnitt, standardavvik og antall observasjoner n (fisker fra 5 år og eldre er ikke med i tabellen).

Tabell 1.

alder	mean	sd	n
2	?	?	3
3	45.2	5.0	9
4	52.2	4.9	50

Oppgave 1 (Med referanse til Tabell 1)

- De 3 fiskene på 2 år hadde lengder på henholdsvis 48, 46 og 35 cm. Finn de manglende tallene i Tabell 1. Test om forventet lengde på 4 år gamle torsker er mer enn 50 cm. Bruk 5% nivå på testen.
- Test på 5% nivå om forventet lengde på 4 år gamle torsker er større enn for 3 år gamle torsker. Hvilke antagelser gjør du?

Oppgave 2

En regresjonsmodell med Y =lengde og X =alder på torsk ble tilpasset i R Commander. Resultatet av kjøringen er gitt i Tabell 2 nedenfor og i form av Figur 1. Bruk resultatene i den grad du finner det nødvendig for å besvare spørsmålene a) – c). (Merk at noen tall er utelatt og erstattet med en stjerne i Tabell 2.)

Tabell 2.

```
Call:
lm.default(formula = lengde ~ alder, data = Fisk)

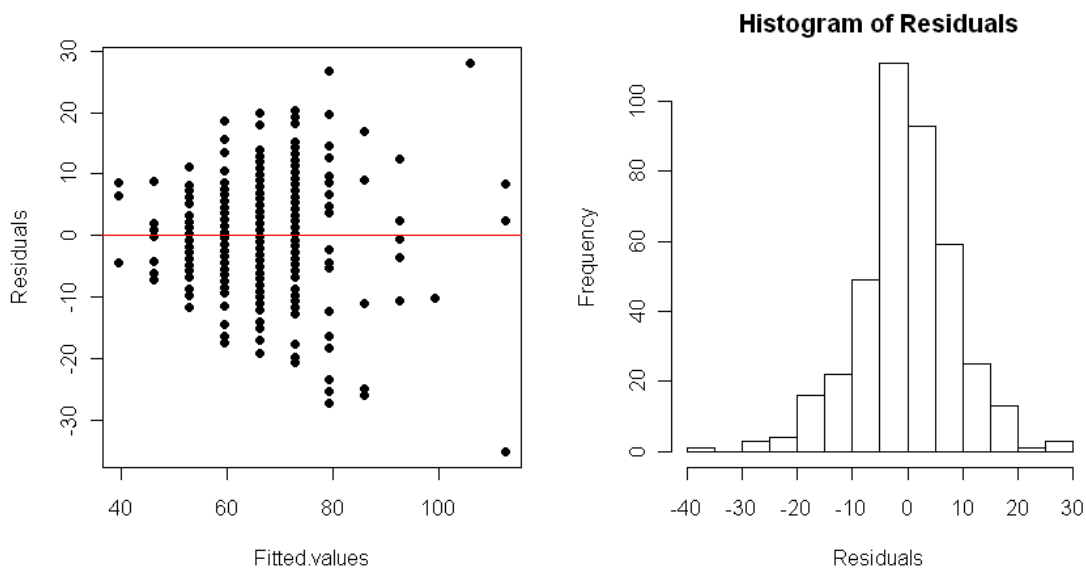
Residuals:
    Min       1Q   Median       3Q      Max
-35.100  -4.437  -0.066   5.240  28.066

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  26.1973     1.7663   14.83    <0.001
alder         6.6447     0.2915     *

Residual standard error: 8.865 on 398 degrees of freedom
Multiple R-squared:  *, Adjusted R-squared: 0.5651
F-statistic: 519.5 on 1 and 398 DF, p-value: *
```

Analysis of Variance Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
alder	1	40828	40828	519.54	< 2.2e-16
Residuals	398	31277	79		
Total	399	*			



Figur 1. Residualer plottet mot tilpassede responsverdier (venstre) og histogram over residualer (høyre)

- Test på 1% nivå om det er en positiv sammenheng mellom alder og lengde på torsker. Finn determinasjonskoeffisienten R^2 og gi en tolkning av denne.
- Hvilke antagelser ligger til grunn for regresjonsmodellen? Vurder om antagelsene ser ut til å være oppfylte.
- Finn et 95% konfidensintervall for forventet lengde på 6 år gammel fisk (Tilleggsopplysning: $\bar{x} = 5.87$)

Oppgave 3

Torskepopulasjonen varierer fra sesong til sesong, og vi skal videre anta at observasjonene gjort i ulike sesonger er uavhengige av hverandre. Nedenfor finner du en R utskrift som kan brukes til å besvare spørsmålene i oppgave 3.

Tabell 3.

Anova Table				
Response: lengde				
	Df	SS	MS	F-value
sesong	(1)	3493	(3)	(5)
Residuals	396	(2)	(4)	
Total	399	72105		
sesong	mean		sd	n
1	68.07		12.1	100
2	67.48		12.7	100
3	64.51		15.6	100
4	60.61		11.9	100

- a) Skriv opp modellen som ligger til grunn for analysen i Tabell 3, fyll inn de manglende verdiene markert med tallene (1) – (5) i ANOVA-tabellen og estimér alle ukjente parametre i modellen.
- b) Sett opp hypoteser for å teste om det er forskjell i forventet fiskelengde mellom sesongene og utfør testen på 5% nivå.
- c) Sesong 1 og 2 er fra to år på 1990-tallet, mens sesong 3 og 4 er fra to år på 2000-tallet. Test om det har skjedd en generell endring i forventet fiskelengde mellom de to tiårene.
- d) Finn et 95% konfidensintervall for forventet forskjell i fiskelengde mellom sesong 3 og 4. Bruk intervallet til å teste om det er signifikant forskjell mellom de to sesongene på 5% testnivå.
- e) Anta at utvalgsgjennomsnittene og utvalgsstandardavvikene for de fire sesongene er som gitt i Tabell 3, men at det bare var 31 observasjoner i hver sesong (altså ikke 100 fra hver sesong). Kan du påvise signifikante forskjeller mellom sesongene på 5% nivå da? Hva med 10% testnivå?

Navn på emneansvarlig: Solve Sæbø

Navn på sensor: John Sølve Tyssedal