

**Bruk av multivariate metoder til å analysere
vekst hos individmerket atlantisk laks
med hensyn på lysregime, kjønn og modning**

Cand. Scient.-oppgave i anvendt statistikk
Retning: Dataanalyse

Kjetil Vestfossen

Matematisk institutt
Universitetet i Bergen



Mars 2005

Takk

Våren 1982 tok jeg statistikkkurset M050 med Dag Tjøstheim som foreleser, mens Hans Karlsen hadde regneøvingen. Både forelesingene, regneøvingen og boken var inspirerende. En glød hos meg for faget var utvilsomt tent.

En takk må rettes til tidligere styrer ved IFM, Ulf Båmstvedt, for at jeg fikk holde kurs i Statistica for Mac for hovedfagsstudentene. Takker også Gunnar Nævdal ved IFM for fine diskusjoner om laksefisk, og for alltid å ha vist meg respekt.

Jeg er takknemlig for å ha fått undervise i statistikk her ved Matamatisk institutt. Å lære vekk faget til intereserte studenter har vært inspirerende, og har gitt faglig styrke.

Ellers takk til alle ansatte ved statistikkgruppen her ved universitetet for imøtekommenhet under studiene. Spesielt vil jeg nevne Ivar Heuch som en inspirator til at jeg tok fatt på et hovedfag i statistikk. Veileder Dag Tjøstheim takkes for stor tålmodighet og for å ha hatt «fokus» på å bli ferdig. Veileder Hans Karlsen var i USA skoleåret 2003/2004, men har hatt gode innspill og en flott humor. Jeg har satt umåtelig pris på friheten til å velge metoder og å løse oppgaven på min måte.

Takker også Vidar Hjellvik for god hjelp til å lage grafer i S-PLUS, samt Mentz Darre Kaarbø for korrekturlesing. En takk må også rettes til Trond Hellem Bø for et år med knallgode og fruktbare kollokvier. Spesielt vil jeg også nevne Karl Ove Hufthammer som jeg har hatt mange gode diskusjoner med om statistikk generelt. Og selvfølgelig en takk til mine medstudenter ellers for sosialt samvær.

Takk også til min sønn Aksel for å ha hjulpet til i hjemmet. Han har måttet gjøre sin del av husarbeidet mens jeg har hatt redusert helse. Etter bruddet i skulderen måtte han både smøre matpakke til meg, kle på meg om morgenen og knytte skolissene.

Jeg tar et hovedfag i statistikk ut fra en fasinasjon for faget, og for å «løse» hovedfagsoppgaven i biologi. Flere års undervisningserfaring i matematikk og statistikk på grunnleggende nivå gjorde et hovedfag i statistikk naturlig.

Rene tilfeldigheter ser også ut til å ha spilt inn.

Innhold

1 Forord	1
1.1 Biologisk motivasjon	1
1.2 Datasettet	3
1.3 Tidligere arbeid med datasettet	5
1.4 Aktuelle metoder	7
1.5 Statistisk motivasjon	9
1.6 Målsettingen	9
1.7 Premisser	12
1.8 Strukturen til oppgaven	13
2 Innledning	16
2.1 Laksens biologi	16
2.1.1 Slektskap	16
2.1.2 Utbredelse	17
2.1.3 Livssyklus	17
2.1.4 Smoltifisering	18
2.1.5 Lysstyring	19
2.1.6 Vekst	20
2.1.7 Atferd	21
2.1.8 Vandringen i havet	21
2.1.9 Miljøfaktorer	23
2.1.10 Kjønnsmodning	23
2.1.11 Livshistoriestrategier	26
2.2 Litt om biologisk tolking	28
2.3 Biologiske problemstillinger	28
2.3.1 Lys	28
2.3.2 Kjønn	29
2.3.3 Modning	29

2.3.4	Klassifisering	29
2.3.5	Vekstkurver	30
2.3.6	Ny måleteknologi	30
3	Om forsøket	32
3.1	Forsøksfisk	32
3.2	Merking	33
3.3	Forsøksoppsett	33
3.4	«Lyskvalitet»	33
3.5	Fôring	34
3.6	Helse	34
3.7	Veiing og måling	34
3.8	SGR og kondisjonsfaktor	36
3.9	Klassifisering i kjønn og modningskategorier	36
3.10	Dataregistrering	37
3.11	Kvalitetsikring av dataene	37
3.12	«Rensking» av uteliggere	37
4	Klassifisering	39
4.1	Klassifisering i hanner og hunner	39
4.2	Klassifisering som postsmoltmodne	39
4.3	Klassifisering som tert	39
4.3.1	Klassifiseringmetoder	41
4.3.2	Grenser i utvalgte arbeider	43
4.3.3	Grenser i vårt datasett	43
4.3.4	Mulige alternative grenser i «vårt» forsøk	45
4.3.5	Diskusjon	47
5	Vekstkurver	51
5.1	Tradisjonelle vekstkurver	51
5.2	Eksponensiell vekst	52
5.3	Termisk vekstkoeffisient	53
5.4	«Powermodellen»	54
5.5	Forsbergs arbeid	55

6 Grafisk framstilling	57
6.1 Fisken ved forsøksstart	57
6.1.1 Lys	57
6.1.2 Kjønn	60
6.1.3 Modning	60
6.2 Ferskvannsfase	60
6.3 Smolt	65
6.4 Samspill: lys og kjønn?	68
6.5 Postsmolt	68
6.6 Vekstforløp	74
7 Statistiske metoder	79
7.1 Valg av statistiske metoder	79
7.1.1 Type studier	79
7.1.2 Konstruksjon av modell	80
7.1.3 Aktuelle metoder	81
7.1.4 Valgte metoder	82
7.1.5 Spesielle problemstillinger	84
7.2 Om estimering og testing	87
7.2.1 Sannsynlighetskvotetesten	88
7.2.2 Multinormalfordelingen	90
7.2.3 MLE og REML	92
7.3 ANCOVA	93
7.3.1 Enveis ANCOVA	93
7.3.2 Variansfunksjoner	94
7.3.3 Toveis ANCOVA	95
7.3.4 Polynomisk regresjon	96
7.4 MANOVA/MANCOVA	97
7.4.1 Hensikt	97
7.4.2 Modeller	98
7.4.3 MANCOVA	100
7.4.4 Kovariansmatrise	101
7.4.5 Testobservatorer	102
7.4.6 En ujustert R^2	103
7.4.7 Ikke-parametriske metoder	104
7.4.8 Foreløpig vurdering	104
7.5 Lineære blandede modeller (LME)	104

7.5.1	Modeller	105
7.5.2	Korrelasjonstrukturer	108
7.5.3	Likelihoodfunksjonen	109
7.5.4	LmList-rutinen	111
7.6	Diggle's modell	111
7.6.1	Modellen	112
7.6.2	Forventningen	112
7.6.3	«Feilleddet»	112
7.6.4	Likelihoodfunksjonen	114
7.6.5	Strukturerte kovariansmatriser	115
7.6.6	Mitt løsnings forslag	116
7.6.7	Variogram	117
7.6.8	Foreløpig konklusjon	118
7.6.9	Lme-rutinen og «Diggles modell»	118
8	Effekt av lysbehandling	120
8.1	Ferskvannsfasen	121
8.1.1	Vekt	122
8.1.2	Lengde	125
8.1.3	Kondisjonfaktor	128
8.1.4	Biologisk diskusjon	133
8.1.5	Vurdering av statistiske metoder	134
9	Kjønn	137
9.1	Ferskvannsfasen	138
9.1.1	Vekt	138
9.2	Sjøvannsfasen	139
9.2.1	Vekt	139
9.3	Diskusjon	141
9.3.1	Biologisk diskusjon	141
9.3.2	Vurdering av statistiske metoder	142
10	Modning	143
10.1	MANOVA	144
10.1.1	Vekt	145
10.1.2	Lengde	148
10.2	MANCOVA	150
10.2.1	Vekt	150

10.2.2	Lengde	152
10.3	Statistikk diskusjon	153
10.4	Biologi diskusjon	155
11	Etterord	158
11.1	Sammendrag av resultater	158
11.1.1	Biologi	158
11.1.2	Statistikk	159
11.2	Perspektiver	160
11.2.1	Datasettet	160
11.2.2	Statistiske metoder	161
11.3	Sluttord	162
11.3.1	Måloppnåelse	162
11.3.2	«Dataverden»	163
A	Lys	172
A.1	Ferskvannsfase	172
B	Kjønn	178
C	Modning	184
C.1	MANOVA	184
C.1.1	Vekt	184
C.1.2	Lengde	185
C.2	MANCOVA	187
C.2.1	Lengde	187
D	LME	189
D.1	Lineær modell	189
D.1.1	Vekt	189
D.2	Modell med tredjegradspolynom	191
D.2.1	Vekt	191

1 Forord

1.1 Biologisk motivasjon

Lakseoppdrett i Norge representerer i dag en næring som eksporterer for ca. 10 milliarder kroner årlig. Det er nesten i samme størrelsesorden som den tradisjonelle norske fiskeri-næring eksporterer for (ca. 15 milliarder kroner ¹). Internasjonalt er også fiske, havbeite og oppdrett av laksefisk en betydelig industri. Kunnskap om laksen er derfor av økonomisk interesse. Laksen har også en verdi i seg selv, som en del av naturen. Sportsfiske har i tillegg en rekreasjonsverdi, og genererer økonomisk omsetning i distriktene. Noen få steder i Norge som i Tana og Alta, er fisket ennå såpass stort at det har karakter av «matauk».

Bestanden av villlaks har gått tilbake over store deler av sitt utbredelseområde på øst- og vestsiden av det nordlige Atlanterhavet (Klemetsen *et al.* 2003). Hovedårsakene har vært knyttet til menneskelige aktiviteter, som forurensning og overfiske. Laksebestanden har naturlige svingninger, men trenden har vært mest nedadgående. Mange enkeltbestander er utryddet eller sterkt svekket. Mulige årsaker til dette kan være:

- Sur nedbør
- Kraftutbygging
- Ødelegging av gytehabitater
- Generell forurensning av elver
- Parasitten *Gyrodactylus salaris*
- Overfiske
- Fiskelus, f.eks. lakselusa *Lepeoptheirus salmonis*
- Rømt oppdrettslaks

¹Mer nøyaktige tall for 2004 er 28,2 milliarder totalt for fiskeri og oppdrett tilsammen. Det ble eksportert oppdrettslaks for ca. 11 milliarder (Blaalid 2005).



Fig. 3. Gammelt stik af Frisenvold laksegård fra slutningen af 1800-tallet. Pælerækken tjener til at holde drivende isflager væk. Fangstkammeret er længst til højre i billedet (fra Illustreret Tidende).

Figur 1.1: Illustrasjon av laksegård (fangstanlegg som stenger av elven), fra Gudenåen i Danmark. Man trodde lenge at laksen i denne elven var utryddet, men den er nå i ferd med å ta seg opp igjen. Dette var tidligere en elv med betydelige fangster. Modifisert etter Jensen (1982).

De fleste av disse årsakene er knyttet til laksens liv i elven. Kunnskapen om laksen i elven er omfattende, men man vet lite om laksens liv i havet. Man er også usikker på om ukjente faktorer spiller inn. På den annen side foregår det et omfattende kultiveringsarbeid for å redde enkeltbestander av laks, samt gjenninnføring i noen elver der man mener laksen er «utdødd».

Da jeg startet på hovedfag i akvakultur i 1989 lå omsetningen i norsk lakseoppdrett på nærmere 6 milliarder kroner. Samtidig var tapet i produksjon grunnet for tidlig kjønnsmodning² anslagsvis 10 prosent. Det vil si at man kunne gjøre besparelser på flere hundre millioner kroner, dersom man kunne redusere dette problemet. Hvorfor noen laks modner etter én vinter i sjøen og andre seinere, var hovedspørsmålet mitt den gang jeg tok til med hovedfaget i akvakultur.

²Dvs. modning som tert = moden etter en vinter i sjøen. En laks med langt fremskreden kjønnsmodning har ingen markedsverdi, fordi mengden protein, fett og farge i kjøttet er sterkt redusert. (Se også kap. 2.)

I dag er problemet med tidlig kjønnsmodning redusert fordi laksen som regel når markedsstørrelse før andre høsten i sjøen. Dette skyldes større smolt ³, bedre fôr og bedre røkting. Avlsprogrammene har selektert for sein kjønnsmodning i mer enn 25 år. Det brukes nå også lysstyring for å få laksen til å utsette modningen. I tillegg sorterer en del produsenter ut den største laksen til slakting for på den måten å redusere tertmodningen. Likevel er tidlig modning fortsatt et problem. Ved forespørsel til en av de største lakseprodusentene i Norge, oppgis et anslag på 3–4 prosent tap pga. tertmodning.

Gjør vi et forsiktig anslag på produksjonstapet til 2–3 prosent pga. tertmodning i 2004, vil dette representere ca. 200–300 millioner kroner. De siste par årene har man observert svært mye tertmodning i enkeltanlegg, ofte opp i 10–20 prosent. Oppdretterne har stilt spørsmål ved om dette kan skyldes den høye sjøtemperaturen de siste årene. Det ser også ut til at man i økende grad registrerer postsmoltmodne ⁴ fisk. Disse vil ofte ikke overleve vinteren i sjøen.

Jeg vil sitere fra *Smolt Kompendiet* fra Felleskjøpet (Wathne 1991, s. 6): «På landsbasis regner man med at omkring 15 prosent av den utsatte smolten i 1990 døde i løpet av de første månedene i sjøen. Skadene var vesentlig større på Vestlandet enn i Nord-Norge. Utsettet i 1990 er estimert til 60 millioner smolt, slik at tapet i antall er 8 til 9 millioner smolt. Verdien av denne smolten er omkring 100 millioner kroner».

Årsaken til det store tapet kan ha vært en kombinasjon av dårlig smoltkvalitet og sykdom. Uansett er både smoltifisering og kjønnsmodning økonomisk viktig innen lakseoppdrett.

Også utfra et ønske om å kunne bevare den naturlige laksebestanden, og som grunnforskning, er studier av laks viktig. Siden laksen er en av de fiskene man har mest kunnskaper om, vil den kunne tjene som kilde til generell biologisk og økologisk kunnskap som også kan øke forståelsen av andre arter.

1.2 Datasettet

Jeg studerte til hovedfag i akvakultur ved Institutt for fiskeri og marinbiologi (IFM) i Bergen i årene 1989 til 1994. Før hovedfaget var ferdig fullført begynte jeg som matematikklærer i videregående skole.

³Dvs. ung laks som kan «settes» i sjøen. (Se kap. 2.).

⁴Dvs. moden første høsten i sjøen.

Den gangen arbeidet jeg med et datasett som inneholdt individmerket laks. Et «elektromagnetisk» merke⁵ ble plassert inne i bukhulen til hver fisk. Merket lot seg lese med en detektor/skanner utenpå fisken. Dette var en relativt ny teknologi i Norge den gang.

Forsøket ble gjennomført i årene 1987–89 på Havforskningens stasjon i Matredal. Røkting av fisken og vekt- og lengdemålinger ble utført av ansatte ved stasjonen. Hovedfagskullet i akvakultur fra 1989 var med på å avslutte forsøket. Jeg var en del av dette kullet. Vi gjorde den siste vekt- og lengdemålingen på fisken, samt utslakting og måling av gonadene⁶. Utfra gonadene ble kjønn og modningsstatus bestemt.

Datasettet bestod i prinsippet av åtte lengde- og vekt målinger per individ, samt at fisken var utsatt for tre ulike lysregimer. Lysbehandlingen foregikk i ferskvannsfasen fra 15. september 1987 til 1 mai 1988. Veksten ble målt også i sjøvannsfasen fra 1. mai 1988 til 27. september 1989, da fisken gikk på naturlig lys (LDN). Man var interessert i effekten av lysbehandlingen på veksten i begge fasene.

Lysbehandlingen var simulert naturlig lysregime (SNL), simulert naturlig lysregime med konstant bakgrunnbelysning (SNL+B) og 24 timers konstant lys (LD24:0).

Hovedmålet med oppgaven var å se på vekstforskjeller 1) hos laks som ble utsatt for ulike lysbehandlinger, og 2) mellom laks som ble kjønnsmoden etter én vinter i sjøen og laks som ble moden seinere.

Veileder den gang ønsket også at jeg skulle finne en «korrekt» statistisk metode til å behandle individmerket fisk, og at dette eventuelt kunne resultere i en manual om hvordan slike data skulle behandles rent statistisk. Opprinnelig var det, slik jeg forstod det, meningen at forsøket skulle være en del av Geir Lasse Tarangers doktorgradsarbeid. Istedenfor ble jeg tilbudt datasettet som hovedfagsoppgave.

Deler av forsøket er tidligere publisert i :

- Stefansson *et al.* (1991)
- Forsberg (1995)
- Vestfossen *et al.* (1991, s. 292)

Ingen av disse publikasjonene utnytter i noen grad de individuelle målingene. Det vil jeg i større grad forsøke å utnytte i denne hovedfagsoppgaven.

⁵Et såkalt PIT-merke som forklares senere.

⁶Dvs. fellesbetegnelse for organer som blir til rogn eller produserer melke.

1.3 Tidligere arbeid med datasettet

Under hovedfaget i akvakultur hadde jeg et kommersielt perspektiv på hva som burde undersøkes. Kjønnforskjeller var jeg ikke opptatt av, fordi dette ikke hadde noen praktisk betydning i kommersielt lakseoppdrett. Jeg ville derimot sammenligne veksten til tidlig og seint moden laks innenfor hvert kjønn.

Videre oppfattet jeg lysregimene nærmest som ekstra støy i forsøket. Lysregimene var ikke kommersielt brukbare, fordi man allerede da visste at de ikke synkroniserte smoltifiseringen. Lysregimenes effekt på veksten i sjøfasen, var til gjengjeld spennende rent biologisk.

Jeg startet dataanalysen med å se på hvert måletidspunkt for seg. På hvert enkelt tidspunkt sammenlignet jeg fisk fra ulike lysbehandlinger ved av hjelp variansanalyse. Avhengig av om data kunne betraktes som normalfordelt eller ikke, brukte jeg en parametrisk eller en ikke-parametrisk metode. En del problemer var knyttet til datasettet og de statistiske metodene. Disse var:

Uteliggere

Datasettet inneholdt mange «uteliggere» som jeg mente kunne komme fra feilregistreringer. Jeg ønsket derfor å finne biologiske yttergrenser for mulige observasjoner, eller eventuelt å «trunkere» materialet. I en samtale med Gunnar Nævdal pekte han på faren ved å kutte interessante deler av materialet. Vi skal se i oppgaven at dette var helt riktig. Jeg slettet fra «arbeidsfilen» observasjoner som etter min oppfatning, var urimelige. Her er det imidlertid rom for diskusjon (se avsnitt 3.12).

Multipel testing

Videre utførte jeg så mange tester at jeg tvilte på om alle forkastningene av hypoteser kunne være riktige. Var det bare tilfeldigheter som gjorde at noen av testene slo ut? Jeg følte den gang at en Bonferroni-korreksjon av signifikansnivået α/n var altfor strengt, og ikke veien å gå (der n er antall tester innen samme familie). Se Wright (1992). Jeg mente imidlertid at å angripe datasettet med mange enkeltstående tester ikke var den beste tilnærming.

For enkle metoder

Dersom resultatene kom tydelig frem ved enkle metoder, mente jeg at det ikke var nødvendig å gjøre noe mer avansert. Da jeg ikke fikk frem tydelige resultater på alle områder, lurte jeg på om det kunne henge sammen med at jeg brukte for enkle metoder.

Manglende data

En ulempe med de mer avanserte metodene var at de dengang kun godtok hele målese-rier. Dersom det manglet en observasjon på en fisk, ble alle målingene «forkastet». Dette var uheldig, særlig sett i lys av at jeg hadde få kjønnsmodne fisker.

Ubalanserte data

Antall modne fisker er mye færre enn antall umodne. Det gikk bra i t-test eller enveis ANOVA, men skulle gi problemer ved toveis ANOVA. Jeg syntes også det var problema-tisk med mange varianter av ettertesting i ANOVA (Høisæter 1989), og at disse generelt hadde lav styrke.

Bimodal fordeling

Parr ⁷ vil ofte utvikle bimodal fordeling. I oppdrett i sjøfasen vil også laksen utvikle bimodal fordeling om den ikke er sortert. Dette kan gi problemer i statistiske metoder som baserer seg på normalfordelingen. I den filen som jeg bruker nå, er nedre modal kuttet ut etter tredje måletidspunkt. Fisk under 100 gram eller 25 cm i desember 1988 er slettet fra filen. Jeg mener bestemt at det ikke var modne fisker blant disse.

Dette gjorde jeg under hovedfaget i akvakultur utfra en delvis biologisk begrunnelse om at disse fiskene er for små til å ta foret. Disse fiskene oppfyller ikke betingelsen føret i overskudd (se avsnitt 3.5). I en ordinær oppdrettsituasjon vil nedre modal rent fysisk være sortert vekk i settefisk-anlegget.

Utelates fisk fra analysen må man skrive at det er gjort, slik som jeg gjorde dengang. Det er mindre heldig å gjøre dette fordi det snevrer inn hvilke fisk som resultane er gyldig for. Det innebærer imidlertid en god del arbeid å finne igjen filene dette ikke er gjort på.

⁷Dvs. liten laksunge i ferskvann. Se kap. 2.

Slettingen hadde altså tre klare konsekvenser, slik jeg så det den gang: 1) Et mer realistisk datasett i relasjon til oppdrett. 2) Innsevret gyldighetsområde for resultatene. 3) Muligheten til å kjøre parametriske tester.

Spesielt vil dette slå ut i mindre varians i MANOVA/MANCOVA, og faktisk mindre «evne» til å skille modne fra umodne i en eventuell logistisk regresjon (og diskriminant-analyse).

Kun ett kar

I sjøvannsfasen går alle fiskene i ett kar. Det kan gjøre tolkingen av resultatene vanskeligere. I ett kar kan man ha kareffekter av ulike slag. Det kan være uregistrert sykdomsutbrudd, uheldig foring i en periode, katter eller mink som besøker karet på jakt etter mat, etc. Jeg antok at dette kun ville virke inn på det samlede vekstmønster og ikke på forholdet mellom ulike grupper av fisk.

1.4 Aktuelle metoder

I manualene til JMP og Statistica leste jeg om repeterte målinger, longitudinelle data og tidsrekker. Jeg mente datasettet hørte hjemme i en av disse kategoriene, men syntes det var vanskelig å forstå metodene fullt ut. Derfor tok jeg kontakt med en statistiker. Han mente at den riktige metoden var logistisk regresjon, i og med at om fisken ble moden eller ikke var en binomisk respons. Jeg hadde ikke tro på metoden, fordi jeg trodde den ville kreve bedre data enn jeg hadde. Men han sa seg villig til å hjelpe med å kjøre en «Growth Model» i MINITAB. Jeg fikk imidlertid ikke noen signifikante resultater og således ikke noe fornuftig ut av kjøringene. Jeg lurte den gang på om dette kunne henge sammen med «for mange uteliggere», at jeg ikke hadde «kvalitetsikret» dataene i relasjon til en kurveproblemstilling eller multivariate metoder (Se av snitt 3.12).

Jeg fikk anbefalt en bok av Snedecor og Cochran (1989) som inneholdt litt om ANCOVA (se nedenfor), og brukte videre denne metoden. ANCOVA er en betydelig forbedret metode i forhold til ANOVA.

MANOVA, MANCOVA, og diskriminantanalyse kunne også være riktig tilnærming. Her kan det være på sin plass å gi noen få stikkord om begrepene.

- ANOVA (Analysis of Variance)

- En kontinuerlig responsvariabel. Kategoriske forklaringsvariabler. Mange grupper kan sammenlignes. Additive effekter av behandlinger eller samspill. Utallige varianter.
- ANCOVA (Analysis of Covariance)
 - En utvidelse av ANOVA. En kontinuerlig responsvariabel. Mange grupper kan sammenlignes. I tillegg til kategoriske også kontinuerlige forklaringsvariabler. Brukes synonymt med multippel regresjon med dummyvariabler.
- MANOVA (Multiple Analysis of Variance)
 - Flere kontinuerlige responsvariabler. Kategoriske forklaringsvariabler. Tar hensyn til «korrelasjonen» mellom responsvariablene ved hjelp av en kovariansmatrise. Multivariat søster til ANOVA.
- MANCOVA (Multiple Analysis of Covariance)
 - Samme som MANOVA bare at det kan være kontinuerlige forklaringsvariabler i tillegg. Multivariat søster til ANCOVA.
- Logistisk regresjon
 - Responsen er en binær variabel eller en kategorivariabel. Forklaringsvariablene er kontinuerlige eller en blanding av kategori- og kontinuerlige variabler. Tar ikke direkte hensyn til «korrelasjonen» mellom forklaringsvariablene. Dette kan gi problemer som er kjent i multippel regresjon (hvis man ikke bruker individeffekter).
- Diskriminantanalyse
 - Responsen er en binær variabel eller en kategorivariabel. Forklaringsvariablene er kontinuerlige eller en blanding av kategori- og kontinuerlige variabler. Tar hensyn til «korrelasjonen» mellom forklaringsvariablene.

Jeg hadde også funnet fram til Diggle (1990), og syntes hans tilnærming i kapittel 5. virket fornuftig. Men jeg visste ikke dengang hvilke program som kunne «kjøre» modellen, og jeg oppfattet også de matematiske utledningene som lite tilgjengelige.

Kort sagt kan man si at jeg hadde nok statistiske kunnskaper til å forstå at å kjøre en ANOVA på hvert tidspunkt ikke var helt korrekt, men ikke nok kunnskaper til å selvstendig velge metode eller til å være sikker på at jeg brukte metodene riktig.

Uansett statistiske kunnskaper, vil en biolog forstå at vekt, lengde og kondisjonsfaktor⁸ «henger sammen» mellom de ulike måletidspunktene. Hvordan dette virker inn på hvilke tester som skal brukes er et annet problem. Det skal vi se grundigere på i denne hovedfagsoppgaven.

1.5 Statistisk motivasjon

Jeg var interessert i å arbeide med metoder som har en vid anvendelse, samtidig ønsket jeg å trekke fordel av min faglige bakgrunn innen biologi generelt og laks spesielt. Å bruke datasettet som jeg hadde fra tidligere, på laks, åpnet derfor muligheten til å kombinere biologi og statistikk. Oppgaven kan sies å grense over til biologi fra statistikk. Men jeg vil vektlegge at dette er en oppgave i statistikk. Multivariate metoder har en vid bruk innen mange fagfelt som biologi, medisin, meteorologi, økonomi og psykologi.

1.6 Målsettingen

Opprinnelig var målet med denne oppgaven:

- Å finne enklest mulige statistiske metoder til å bruke på datasettet som jeg har.
- Å implementere metodene i Splus.
- Å gjøre tilpasninger av modellene.
- Å evaluere metodene.

Planen var å simulere data med lignende strukturer etter å ha kartlagt datasettet. Dette for å se på styrke og robusthet til de ulike metodene. I utgangspunktet gjorde det ikke noe om man beholdt en del «uteliggere». Da kunne man etterpå studere deres innflytelse i modellene.

Jeg hadde for vel 10 år siden sett på en modell beskrevet av Diggle (1990) og syntes hans tilnærming virket fornuftig i forhold til datasettet. Etter å ha forstått modellen til Diggle (1990) bedre så jeg at å koble individuelle datamerker⁹ med rammer¹⁰ som måler fisken

⁸Kondisjonsfaktor er et mål på forholdet mellom lengde og vekt. Se 3.8.

⁹Datamerker er her PIT-merker (Passiv Integrator Transponder) beskrevet i avsnitt 3.2.

¹⁰Rammene kan «fotografere» fisken når den svømmer gjennom, slik at man får da et lengde- og vektestimater.

automatisk under vann, virkelig hadde noe for seg. Den tilnærming som Diggle (1990) anbefaler var dermed aktuell å studere nærmere. Målsettingen til denne oppgaven er derfor todelt.

- Å finne enklest mulige statistiske metoder til å bruke på lignende datasett som det jeg har dvs. med gjentatte målinger over tid på samme individ.
- Å se teoretisk på hvilke metoder som er aktuelle ved «ny måleteknologi» (se innledningen) fortrinnsvis modellen til Diggle (1990).

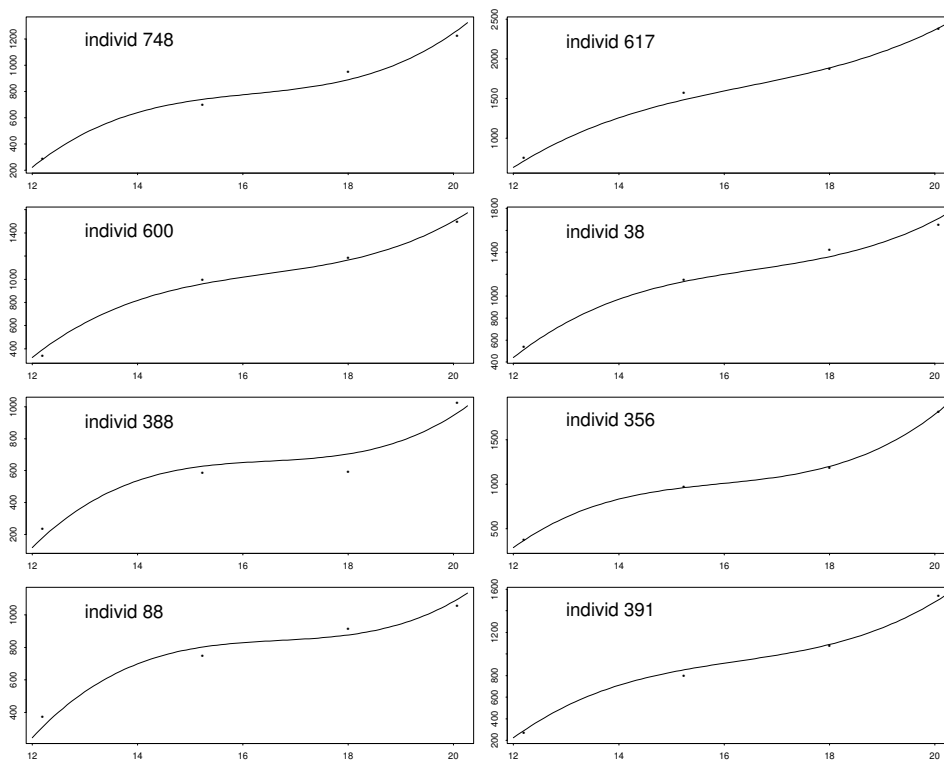
Målet har forandret seg flere ganger, men underveis var det meningen å estimere en passelig vekstkurve ved hjelp av nlme-rutinen (se kap. 7) som står for «Nonlinear Mixed-Effects» (Pinheiro og Bates 2000).

Mens jeg prøvde å overparametrisere en modell med et tredjegradspolynom med individ-effekter i lme-rutinen ¹¹, så jeg at dette tilsynelatende fungerte svært bra. Jeg har derfor benyttet dette som en av metodene. (Se figur 1.2 på neste side som et eksempel på individuelle tilpasninger av modellen.) Men det viste seg at denne metoden ikke tilfører mer kunnskap om hovedproblemstillingen rent biologisk enn MANOVA og MANCOVA. Derfor viser jeg selve bruken av denne metoden i appendix D. Men jeg mener denne metoden representerer en god måte til å angripe vekstforsøk, spesielt ved relativt lange tidserier.

Det meste av de statistiske metoder som jeg bruker vil fremstå som trivielle for en erfaren statistiker. Jeg forslår imidlertid to litt spesielle metoder som jeg ikke har funnet omtalt i den litteratur som jeg har brukt. Det første er jeg gjør omfattende bruk av individeffekter for å forklare variansen, og for å gjøre de individuelle vekstkurvene mest mulig fleksible (nevnt overfor). Det andre er at jeg foreslår en annen variabel enn tid til bruk for å parametrisere kovariansmatrisen (se avsnitt 7.6.6).

Tidsrekker, vekstkurver, MANOVA, MANCOVA og logistisk regresjon er også som nevnt tidligere aktuelle metoder ved vekstforsøk med individmerket fisk (se 1.4 og kap.7). Det kan imidlertid se ut som metoder for å behandle individmerket fisk, oppfattes som vanskelige og utilgjengelige. Derfor kunne det være et mål å gjøre rede for noen av disse metodene på en enkel måte. Det gjennomfører jeg kun i den forstand at jeg beskriver også enklere modeller enn dem jeg bruker.

¹¹«Linear Mixed-Effects»



Figur 1.2: Individuelle plott (Vekt4-6). (Individer er valgt ved «random» funksjon). Modell: tredjegradspolynom med modning nøstet innen kjønn. (Se kap. 7 og 10 for flere opplysninger).

1.7 Premisser

Jeg har stått fullstendig fritt med hensyn hvordan jeg har villet løse oppgaven. Det eneste kravet har vært at jeg skulle bruke Splus og skrive i L^AT_EX. Dersom datasettet var for dårlig var planen å simulere data med lignende strukturer, og likevel bruke multivariable teknikker.

Jeg har selv valgt å begrense meg til å angripe datasettet med standardiserte metoder som finnes som ferdige rutiner i Splus. Kun når det har vært strengt nødvendig har jeg transformert data. Jeg har også unngått å bruke bootstrapping¹² og EM-algoritmen¹³. Målet mitt har vært å finne enklest mulige metoder som likevel er tilfredsstillende (og lett kan brukes av andre enn statistikere).

Videre har jeg forsøkt å unngå å fordype meg i de samme tema som Imsland (2001). Vekst generelt hos fisk, vekstkurver og utledning av bruk av F-fordelingen på multivariate data, er grundig diskutert hos ham.

Jeg tar i minst mulig grad opp de samme problemstillingene som Stefansson *et al.* (1991). De drøfter grundig lys som faktor sammen med en del smoltfysiologi, samt lysets påvirkning på antall modne, og da trenger ikke jeg gjøre det. Samtidig gjør jeg oppmerksom på at de har fem måletidspunkt i ferskvannsfasen, mens jeg bare har to av disse. Dette medfører at jeg analyserer denne fasen i minst mulig grad.

Opgaven nå handler om statistikk, men dette bør ikke sees isolert fra fra problemstillingene som er biologiske. Jeg er skeptisk til å bruke statistikk (på et datasett) som ikke er drevet av teori, fordi det da blir «data-mining» eller fising i dataene (se kapittel 7).

Intensjonen er ikke å gi datasettet en fullstendig biologisk tolking, men heller å gå inn i noen problemstillinger og å se på statistiske metoder knyttet til disse.

Det var og enighet om å bruke de mest «oppdaterte filer» jeg hadde av datasettet med hensyn til utrensning av «uteliggere». Disse var:

- vlk.mindre.stat
- vlsgr.mindre.stat
- rå.gonader.xls

¹²Metoder for reestimering. Se avsnitt 7.1.5.

¹³Metoder for rekonstruksjon av manglende data. Se avsnitt 7.1.5.

Filene var organisert med hver fisk på en rekke. Filen «vlk.mindre.stat» inneholder åtte variabler (kronologisk) for vekt, lengde og kondisjonsfaktor, mens filen «vls-gr.mindre.stat» er organisert på samme måte bare at kondisjonsfaktor er byttet ut med «SGR»¹⁴ for vekt. Filene inneholder også koder for lysregime, kjønn og modingsstatus. For å kunne benytte meg av lme- og nlme-rutinen måtte filene organiseres med vektene og lengdene til alle fiskene i to variabler og en egen tidsvariabel, samt variabler for individ etc.

På grunn av begrensinger i filstørrelse (et maks antall variabler) i Statistica, inneholder ikke disse to første filene PIT-numrene¹⁵. Det er derfor vanskelig å koble disse til den siste filen som inneholder PIT-numrene, vekt, lengde, gonadevekter og GSI¹⁶ fra siste måletidspunkt. Filene med alle vektene og lengden på fiskene lar seg ikke koble til filen innholdene «gonadene», derfor brukes den opprinnelige oppdeling i modne og umodne. (Jeg forsøkte å koble («matche»)¹⁷ de på grunnlag av siste vekt og lengde, men fikk da endel fisk uten entydig «match».) Vi beholder dermed den inndeling i modne og umodne som er gjort opprinnelig.

Jeg har også forsøkt å se på datasettet (se neste avsnitt) med et åpent sinn, og minst mulig grad se på gamle resultateter (med unntak av «Material og metode» kapitlet). Først mot slutten av denne oppgaven har jeg sett mer grundig på det gamle utkastet til hovedoppgave. Det er nesten 10 år siden jeg arbeidet med hovedoppgaven i biologi, slik at jeg derfor i noen tilfeller må rekonstruere hva som var gjort dengang. Jeg ikke er heller ikke oppdatert på alle biologiske problemstillinger. Jeg har vektlagt å finne relevante statistiske metoder og å gjøre tilpasninger av metodene til datasettet, ikke å gjøre en fullstendig biologisk tolking.

1.8 Strukturen til oppgaven

Jeg følger ikke IMRAD-formen (Introduction, Materials and methods, Results and Discussion), men velger heller å i hovedsak behandle hver problemstilling mest mulig ferdig innenfor hvert kapittel.

¹⁴Prosentvis vekstrate. Defineres ulikt hos ulike forfattere, se derfor avsnitt 3.8.

¹⁵PIT-nummer er en 12 sifret alfanumerisk kode til å identifisere hver enkelt fisk ut i fra de individuelle merkene (se kap 3.)

¹⁶Dvs. gonadevekt delt på vekten til hele fisken. Se kap. 5

¹⁷«match» = rutine i en del dataprogrammer for plasere variabler sammen på grunnlag av entydig identifikasjonøkkel.

I kapittel 2 skriver jeg litt generelt om laksens biologi, for å gi andre enn fagfolk på fisk (for eksempel en statistiker) mulighet til å sette seg bedre inn i de biologiske problemstillingene.

Kapittel 3 gir utfyllende opplysninger om forsøket, og jeg ser jeg litt på metoden for å klassifisere laksen i ulike modningskategorier i kapittel 4. Videre har jeg skrevet kort om vekstkurver innen akvakultur i kapittel 5.

De samme statistiske metodene brukes innenfor flere kapitler, og derfor har jeg et generelt kapittel om statistiske metoder. Det er ikke lett å unngå en del matematikk og spesielt lineær algebra i fremstilling av metodene, men jeg håper likevel at biologer også kan ha glede av å lese deler av oppgaven.

Grafisk framstilling i kapittel 6 kommer før statistiske metoder i kapittel 7. Dette for å gi innsikt i motivasjonen til valget av metodene. Jeg har også sett på modellen beskrevet av Diggle (1990), og vurdert på muligheten av å bruke hans modeller i forbindelse med ny måleteknologi¹⁸ Derfor ser jeg på hans modell i kapitlet om de statistiske metodene.

Faktorene lys, kjønn og modning henger sammen, slik at disse kapitlene (8, 9 og 10) ikke kan sees isolert fra hverandre. Siden kjønn viser seg å ha liten effekt på vekstmønsteret, ser jeg på lys først og deretter mest på modning innen kjønnene.

Bare rundt to tredeler av fisken «overlever» hele forsøksperioden og er kjønnbestemt og vurdert som moden eller ei. Ved å kutte ut kjønn- og modningsinformasjonen kan vi da bruke mange flere fisk når vi ser på effekten av lys alene.

Jeg har plassert sammendraget av resultatene i et etterord, som også tar opp litt perspektiver fremover.

Alternativ struktur

Jeg kunne sett først på effekten av lys. Deretter lys og kjønn. Til slutt kunne jeg sett på lys, kjønn og modning sammen.

Dette ville da kunne gi resultater som kan sammenlignes med andre eksperimenter. I de fleste lyseksperimenter har man ingen opplysninger om kjønn og modning. I eksperimenter med kjønn og vekst, kan forsøket være avsluttet før man vet noe om eventuell modning. Vi får da en trinnvis avdekking av hva som «skjuler» seg i dataene. Hadde vi

¹⁸Ny måleteknologi som nevnt tidligere og som innebærer at fisken blir målt i vannet, dersom man bruker rammer som «fotograferer» fisken fra flere vinkler når fisken svømmer gjennom.

hatt informasjon om familiebakgrunn til fiskene ville dette blitt enda ett trinn i prosessen.

2 Innledning

2.1 Laksens biologi

2.1.1 Slektskap

Atlantisk laks (*Salmo salar* L.) tilhører familien *Salmonidae* (laksefisker), som består av slektene *Salmo*, *Salvelinus* (røye, bekkerøye, med flere) og *Oncorhynchus* (Stillhavslaks). Slekten *Salmo* består av to arter: Atlantisk laks og ørret (*Salmo trutta*). Regnbueørreten (*Oncorhynchus mykiss*) ble tidligere regnet til slekten *Salmo* med navnet *Salmo gairdneri* (Hansen 1998, s. 17). Artsbegrepet, slik det vanligvis brukes, kan ikke alltid anvendes på fisk, da f.eks laks og ørret gir fertilt avkom.

Laksefiskene har en fettfinne mellom ryggfinnen og halen. Familien antas å ha opprinnelse fra en felles ferskvannsfisk, men dette er omstridt (Hansen 1998, s. 17–18). Laksen i Østersjøen (baltisk laks) regnes som en egen bestand.

Laksen kalles en anadrom fisk fordi den gyter i ferskvann, men har vandring til saltvann. (Ålen er et eksempel på en katadrom fisk som gyter i saltvann, men har opphold i ferskvann). Laksefiskene er i varierende grad anadrome. Med det mener man at det er forskjeller i tiden de oppholder seg i ferskvann og saltvann. Mange bestander vandrer ikke ut i havet i det hele tatt. Donaulaksen vandrer ikke ut i havet, og regnes ofte som en egen art (*Hucho*).

I Norge har vi også bestander av laks som ikke vandrer ut i havet (relikte): småblanken overfor Fiskumfoss i Namsen og blegen i Byglandfjorden (Wathne 1991). Disse bestandene er fullstendig genetisk isolert. Det fantes også en tilsvarende bestand i Nidelva i Aust-Agder, men den antas utryddet (Schandy og Røed 2002, s. 15).

2.1.2 Utbredelse

Atlantisk laks er naturlig utbredt på begge sider av det nordlige Atlanterhavet og i Den Gamle Verden. Ørreten er utbredt på det europeiske kontinent, på Island, Færøyene, Svalbard og på de britiske øyer. Slekten *Salvelinus* har en cirkumpolar utbredelse, mens slekten *Oncorhynchus* hører hjemme på begge sider av det nordlige Stillehavet.

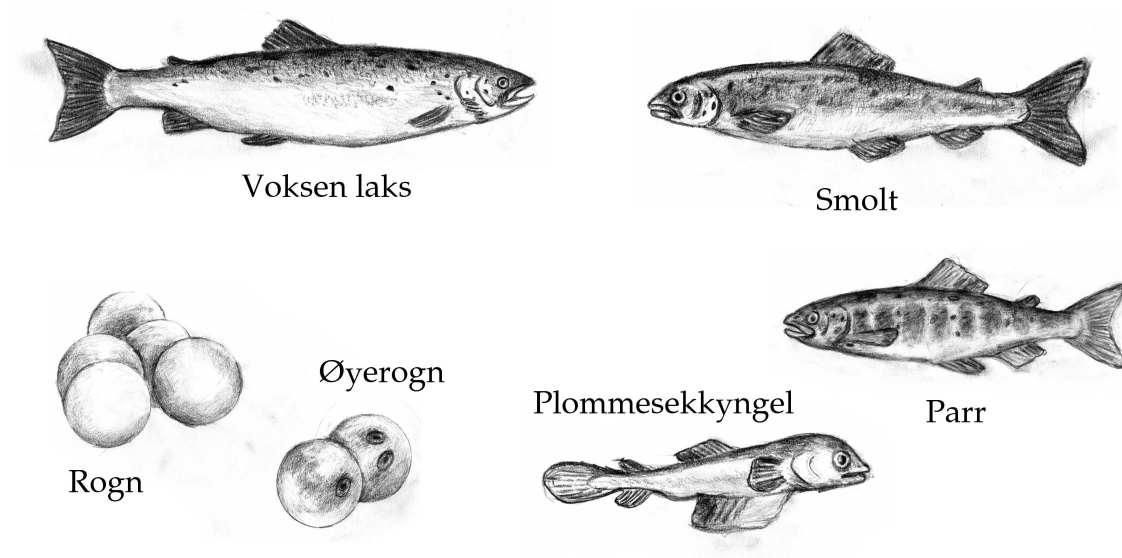
Spesielt regnbueørreten har fått en mye videre utbredelse gjennom kunstig utsetting i naturen, og gjennom oppdrett. Atlantisk laks har muligens fått en videre utbredelse gjennom rømming fra oppdrett, blant annet i Chile og på vestkysten av Nord-Amerika (Schandy og Røed 2002, s. 15).

2.1.3 Livssyklus

Norsk Atlantisk laks gyter i elver seinhøstes eller tidlig vinter. Eggene legges nedgravd i en blanding av grov grus og småstein på bunnen av elven. Eggene klekkes påfølgende vår, etter ca. 400–480 døgngrader (temperatur*antall døgn) (Ingebrigtsen 1982, s. 99). Rett etter klekking kalles den plommeseckyngel, fordi den har en sekk med plommemasse hengende under buken. Når plommesekken er fortært, kalles yngelen parr. Man sier da at den har elvedrakt, med blant annet karakteristiske striper på siden (kamouflasje som passer til elvehabitatet, se figur 2.1 på neste side).

Utvandringen til havet skjer om våren/forsommer ca. ett til åtte år etter klekking. Antall år til den vandrer ut ser ut til å avhenge av vekstbetingelsene. Vi sier at den smoltifiserer (se nedenfor) før den vandrer ut i havet. Etter et varierende antall år i havet på næringsøk vender den tilbake til den elven den har vokst opp i for å gyte. Den kan da sies å ha fullført livssyklusen.

Etter gytingen vil laksen vandre ut i havet igjen. Tidspunktet varierer fra få dager til flere måneder etter gyting. Overlever gytelaksen, vender den tilbake til elven for å gyte igjen, oftest etter ett eller to år i sjøen. Dersom det går flere år til neste gyting, betegnes de mellomliggende år for hvileår. Typisk vil mindre enn 10 prosent av den anadrome laksen gyte en andre gang, enda færre en tredje gang (Fleming 1996). Atlantisk laks er altså en flergangsgyter (iteropar). Dette i motsetning til laks i Stillehavet som i regelen er engangsgyter (semelpar).



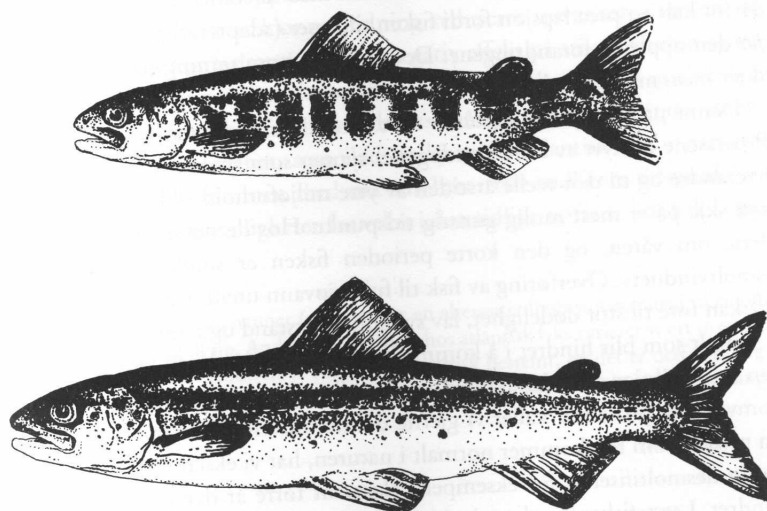
Figur 2.1: Illustrasjon av ulike stadier i livssyklusen til laksen. Relative størrelser er ikke riktige. Rogn (nederst til venstre), øyerogn, plommeseekkyngel, parr og smolt (øverst til høyre) og voksen laks (gjeldfisk, dvs. ikke ytre tegn på å være kjønnsmoden). Illustratør: Jan Henning Olsen.

2.1.4 Smoltifisering

I månedene før laksen vandrer ut i havet skjer det en morfologisk, fysiologisk og atferdsmessig tilpasning til sjøvann. Denne forandringen kalles smoltifisering. Se f.eks. Saunders *et al.* (1994) og referanser i denne. Tilpasningen til sjøvann skjer altså før den vandrer ut i havet. Under smoltifiseringen skifter laksen farge til en sølvblank fisk med mørke render, blant annet på halefinnen. Også fasongen forandres ved at fisken blir mer slank og langstrakt. Spesielt vil partiet foran selve halefinnen strekke seg. (Se figur 2.2.)

Smoltifiseringen inntreffer vanligvis om våren eller forsommeren i vill tilstand. Daglengden regulerer når på året smoltifiseringen inntreffer. Temperaturen i vannet kan fremskynde eller forsinke prosessen. I naturen finner man best sammenheng mellom hastigheten på temperaturøkningen og utvandringstidspunktet (Wathne 1991, s. 11).

Den store variasjonen i smoltalder i naturen antas i hovedsak å skyldes forskjeller i næringstilgang og vekstbetingelser. Parr i oppdrett som ikke skal smoltifisere påfølgende vår har redusert vekst gjennom vinteren, uten at man vet helt sikkert hvorfor. Det fins indikasjoner på at parren må nå en minste størrelse (8–10 cm etter endt vekstsesong, men



Figur 2.2: Illustrasjon av parr (øverst) og smolt (nederst). Modifisert etter Hansen (1998).

det kan nok være mer komplisert enn bare lengde) før den starter smoltifiseringprosessen (Wathne 1991, s. 17). Ofte vil det i oppdrett utvikles en bimodal fordeling på lengde i løpet av høst/vinter (Skilbrei 1991). En mulig tolking er at parren i nederste modal har «bestemt» (eller har som strategi) å smoltifisere først et år senere enn øvre modal.

Fysiologisk skjer det forandringer slik at fisken er i stand til å tåle saltvann. Se Hoar (1988) og referanser i denne. I gjellene øker antallet av kloridceller, og disse cellene blir større, endrer form, og ligner mer typiske kloridceller hos saltvannsfisk. Samtidig øker Na-K-ATPase-aktiviteten i cellemembranene. Sammensetningen av fettsyrer og proteiner i kroppsvev endres, og metabolismen øker (Wathne 1991, s. 22–23).

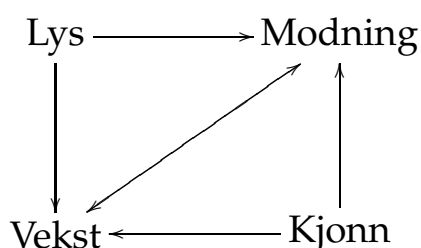
Genetiske forskjeller innen og mellom elvestammer og familier (her: søskengrupper) antas også å virke inn på alder ved smoltifisering. Undersøkelser om hvorvidt eldre smolt modner tidligere enn yngre smolt har gitt lite konsistente resultater. Man kalte da dette teorien om negativ korrelasjon mellom smoltalder og sjøalder (Nævdal *et al.* 1979). (Se avsnitt 2.1.10 om kjønnsmodning for utfyllende opplysninger.)

2.1.5 Lysstyring

Laksen antas å ha en form for indre rytme som fungerer som en slags klokke. Den har både en døgnrytme (*circadian*) og en årsrytme (*circannual*). Årsrytmen kan til en viss grad

overstyres ved å regulere daglengden over en periode (lysregime). Man kan på denne måten simulere årstidene. Daglengden kalles da «Zeitgeber», eller på norsk «tidsgiver» (Lærum 1988).

I dagens settefiskanlegg styres smoltifiseringen ved å manipulere lysperioden. Anslagsvis en tredel av produksjonen er nå høstsmolt (halvårs-smolt). For å få dette til simulerer man en vinter og en vår i løpet av sommeren, og utpå høsten er parren smoltifisert. Tidspunktet for når laksen gyter innen en sesong kan også styres. Ved å simulere en for-



Figur 2.3: Enkel modell av forholdet mellom faktorene i eksperimentet. Tre faktorer påvirker modningen, som igjen påvirker veksten. Pilene angir årsaksretning.

kortet eller forlenget årsrytme vil man henholdsvis fremskynde eller forsinke gytingen. Det er også rapportert at man kan utsette kjønnsmodning en sesong ved å sette på ekstra lys tidlig på vinteren (Oppedal 2002). Men man kan også få motsatt effekt (Kråkenes *et al.* 1991), avhengig av når ekstra lys gis. Figur 2.3 viser en enkel modell for de nevnte forsøk, samt det forsøk jeg ser på i denne oppgaven.

2.1.6 Vekst

Veksthastigheten til laksen er delvis genetisk bestemt, men den påvirkes også av ytre faktorer. Ytre og indre faktorer påvirker veksten gjennom ulike former for samspill. Se f.eks. Imsland (2001) og referanser i denne.

Veksten vil avhenge av temperatur, daglengde, årstid og livsfase til fisken. Med livsfase mener jeg her for eksempel smoltifisering og kjønnsmodning.

Vektøkningen i slutten av smoltifiseringsfasen er ikke nødvendigvis så rask, mens lengden øker raskt i denne fasen. En stor andel av energien går da med til fysiologiske og anatomiske forandringer hos fisken.

Videre er det i oppdrett observert høyere vekst om våren hos laks som skal gyte påfølgende høst, mens denne veksten igjen dempes i løpet av sommeren og stopper mer eller

mindre helt opp etterhvert som det nærmer seg gyting. Man antar at villaksen slutter å spise når den går opp i elven.

Veksthastigheten (regnet som SGR¹) er også svært forskjellig i ferskvanns- og sjøvannsfasen (i startfôringsfasen kan den prosentvise daglige tilvekst være opp i 5 prosent, mens på en større laks i sjøfasen er 0,5 prosent mye). Dette vil ofte også henge sammen med temperatur og mattilgang. Man kan si den vokser svært godt i sjøen (i absolutt verdier), både i oppdrett og i vill tilstand. Laksen ser også ut å ha en evne til kompensasjonsvekst dersom den i perioder har liten tilgang på føde (Metcalf og Monaghan 2001).

2.1.7 Atferd

Man kan tenke seg at laksen har en atferd som forsøker å optimalisere inntaket av energi samtidig som den vil unngå predatorer og aggressive konkurrenter.

Vill parr har territorial atferd i elven. Både i naturen og i oppdrett er det påvist sosial rangordning i ferskvannsfasen. Når laksen smoltifiserer, vil den etterhvert stå høyere i vannmassen, og det dannes grupper av fisk som vandrer ut av elven sammen. Disse gruppene vil muligens holde sammen en stund etter utvandringen.

I oppdrett vil også veksten i sjøvannsfasen være avhengig av sosial rangordning i merden (not som stenger fisken inne) eller eventuelt i karet ved landbasert oppdrett.

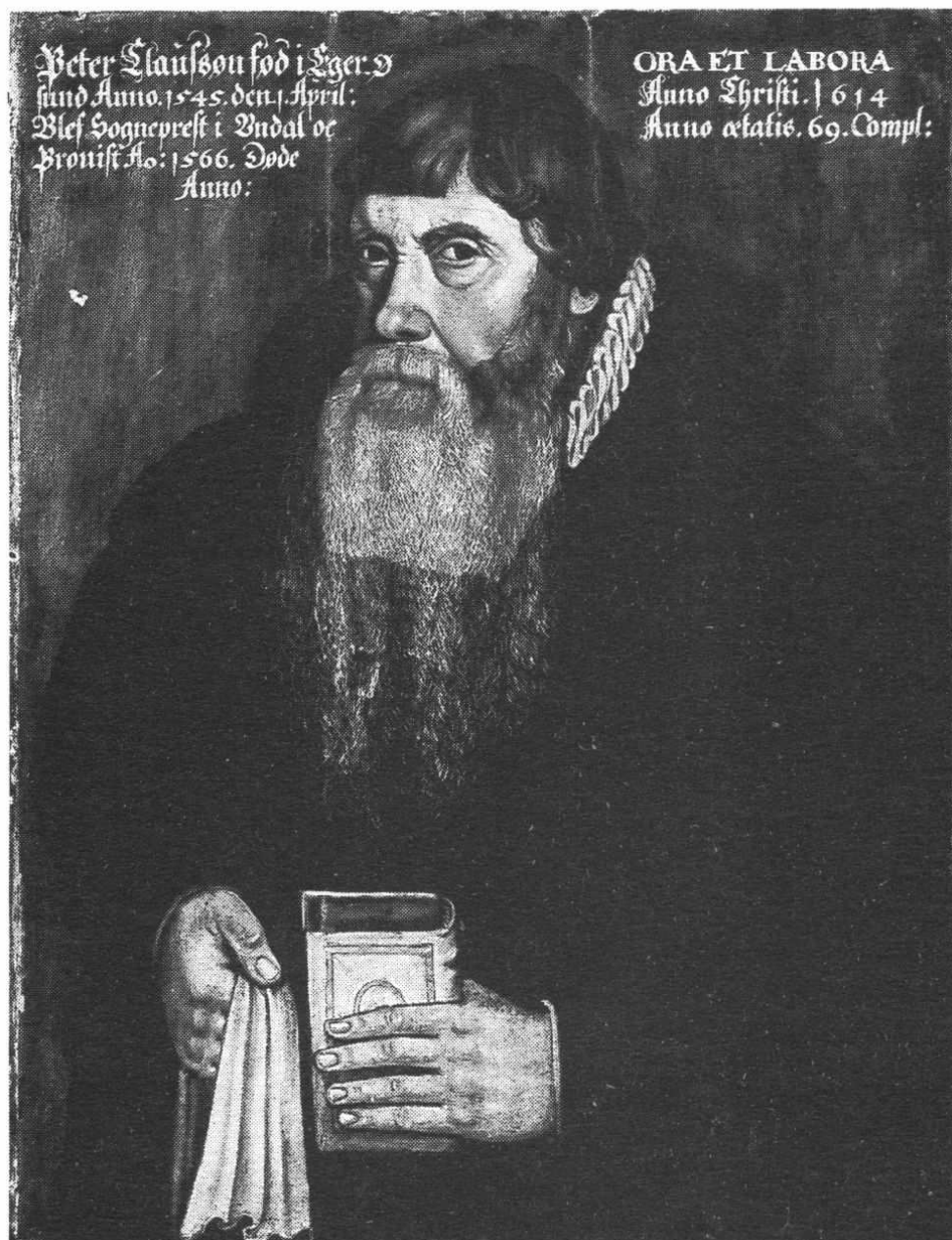
Under gytingen vil hannene konkurrere om hunnene. Hannene er gyteklare over en lengre periode enn hunnene, slik at det til enhver tid er flere «klare» hanner enn hunner tilstede (Fleming 1996).

2.1.8 Vandringen i havet

Laksen oppholder seg ikke lenge ved elvemunningen, og vandrer relativt fort ut av fjordsystemet. Norsk laks kan vandre så langt som til vestkysten av Grønland, men i regelen vil den oppholde seg i Norskehavet eller i Barentshavet. I havet oppholder laksen seg nær overflaten, på søk etter byttedyr som småfisk og ulike reke- og krepsearter (Schandy og Røed 2002).

Merkeforsøk viser at ved tilbakevandring til kysten kan laksen komme inn et godt stykke fra den riktige elven. Den følger så kysten mot den elven den er oppvokst i (eller det stedet de er sluppet ut fra ved havbeite). En del laks vil også vandre opp i «feil» elv.

¹SGR = daglig prosentvis tilvekst, og er definert i avsnitt 3.8.



Peder Claussøn Friis. Foto O. Væring.

Figur 2.4: Illustrasjon av Peder Claussøn Friis (1545–1614), som skrev at laksen vender hjem til den elv den er født i. Han mente også at «Laksestørjen ledet laksen, men Gud ledet laksestørjen». Modifisert etter Berg (1986).

En teori om laksens navigering har gått på at smolten legger igjen ett luktspor i sjøen som gytelaksen følger på vei til elven («feromonhypotesen»). Nyere forskning forkaster «feromonhypotesen», og støtter heller at laksen «lærer veien» under utvandring som smolt (Schandy og Røed 2002, s. 30–34). Mye laks kommer visstnok inn utenfor Hitra og Frøya.

2.1.9 Miljøfaktorer

Det er gjort utallige undersøkelser der man forsøker å relatere alder ved smoltifisering og modning til ulike miljøfaktorer. Man finner for eksempel en tendens til økende smoltalder jo lengre nord elven ligger. Årsaken antas å være kaldere vann (lavere metabolisme) og dårligere vekstvilkår i nordlige områder.

Man har også forsøkt å relatere alder ved modning til kombinasjonen sjøtemperatur, lengdegrad, vannmengde i elven og elvelengde (Scarnecchia 1983). Men det kan se som man ikke tar hensyn til genetiske forhold.

Det er observert en tendens til at bestander i elver med lav vintertemperatur gyter tidligere i sesongen (Heggberget 1988). Dette kan å være en tilpasning eller «timing» til den lokale elvens «vårløsning», dvs. at mattilgangen i form av insektlarver og smådyr skal være god når plommesekken er fortært (Fleming 1996).

2.1.10 Kjønnsmodning

På samme måte som smoltifiseringen kan inntreffe etter et varierende antall år i ferskvann, vil gytingen første gang skje etter et varierende antall år i sjøen (Alm (1959), Gardner (1976), Fleming (1996) og Klemetsen *et al.* (2003)). Mønsteret (variabel sjøalder) kan være forskjellig for ulike elver og for hanner og hunner.

Hanner som ikke vandrer ut i sjøen før de blir modne kalles dverghanner eller snikere (og på engelsk «sneakfuckers»). En smolt (som regel en hann) som gyter etter kun en sommer i sjøen kalles postsmoltmoden.

Som oftest oppholder laksen seg i sjøen mellom en og tre vintre før den skal gyte første gang. En laks som kommer tilbake etter en vinter i sjøen kalles tert (1SW²), etter to år mellomlaks (2SW) og etter flere vintre storlaks. Det er faktisk observert laks som er førstegangsgytere når de er eldre enn tre sjøvintre. Hannene er ofte tert, mens hunnene er

²1SW = one sea winter

mellomlaks når de gyter første gang, uten at dette er noe fast mønster. Generelt ser ut til at hunnene modner senere enn hannene (Thorpe 1994).

De første gytelaksene ankommer norskekysten i april/mai, mens de fleste ankommer elven mellom mai og august (Schandy og Røed 2002, s. 30–31). Storlaksen ankommer ofte tidligst, uten at dette er noe fast regel i alle elver. Norsk laks gyter hovedsakelig i perioden oktober til januar (Heggberget 1988).

Kjønnsmodning har jeg ikke sett definert. Men mitt forslag er at man må knytte dette til at gonadene starter å vokse, fysiologiske forandringer eller atferdsmessige forandringer. Dette må da stå i motsetning til laks som ikke skal gyte samme sesong. Vi kan tenke oss at i det laksen starter vandringen i havet mot elven for å gyte så må «beslutningen» være tatt.

Kjønnsmodningen til laks i oppdrett som skal bli 1SW (tert) antas å starte omtrent ett år før selve gytingen (Youngson *et al.* (1988), Thorpe (1989) og Taranger (1993)). Frigjøring av hormonet gonadotropin (1) kan være et av de første sikre tegnene på kjønnsmodning (Olsen (1991) og Taranger (1993)). Det er også funnet hormoner som frigjør gonadotropin (1). Siden disse hormonene ikke er målt på fisken i dette forsøket, går jeg ikke grundigere inn i det her og nå.

Det er videre ikke målt hormoner på laks fanget i havet, såvidt jeg vet. Jeg antar at frigjøring av hormonet gonadotropin(1) også eventuelt kan utløse selve tilbakevandringen. Samtidig som gyte-vandringen starter tas kanskje også en endelig «beslutning» om å gyte denne sesongen. Dette kan dog ikke være tilfelle for de laksebestander, der laksen noen ganger står et helt år i elven før den skal gyte.

Om veksten påvirkes av modningen eller motsatt er et vesentlig biologisk spørsmål i denne oppgaven. For eksempel finner Hunt *et al.* (1982) forhøyet vekst hos modnende tert-hanner vinter/vår og topper av nivået av 11-oxotestosterone i samme periode. Her er det mulig å anta at modningen medfører forhøyet vekst og ikke motsatt (Taranger 1993).

Mange har postulert teorier om at laksen fremskynder kjønnsmodningen (én sesong) ved gode vekstbetingelser, eller motsatt at den utsetter modningen ved dårlige vekstbetingelser (Rowe *et al.* (1991) og Thorpe (1994)). I et forsøk beskrevet i Thorpe *et al.* (1990) ved sulting av laks i februar–mars hos hanner, og i mars–april hos hunnene, oppnåes en redusert andel modning. Dette støtter opp om at «beslutningen» for en potensiell tert om å modne eller ikke modne kommende høst, tas på vinteren eller våren.

Det er også foreslått at modningen kan slås av etter at den er startet, helt fram til utpå våren. Man har da en «to-stegs beslutningsmodell» (Taranger 1993, s. 31). Figur 2.3 på side 20 og figur 2.5 på neste side viser enklere modeller enn «to-stegs beslutningsmodellen».

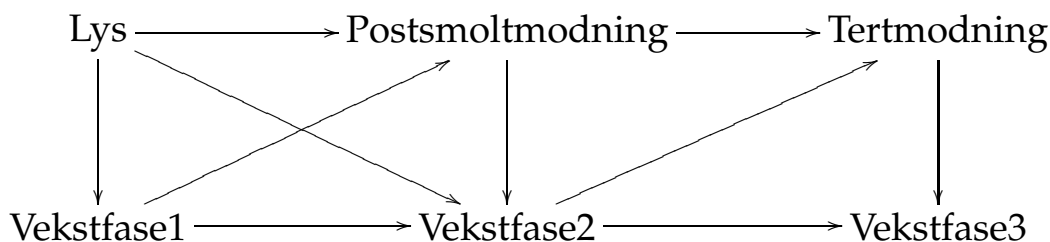
Arv vil også virke inn på alder ved kjønnsmodning. Man tror det er tilpasninger til forhold i elvene, og det er utført forsøk med ulike stammer og familier i oppdrett (Nævdal *et al.* (1978a) og Gjerde (1986)).

Også ut fra en økologisk tankegang må laksen ha en viss mengde fettreserver for å kunne klare å overleve en energikrevende gyting (Thorpe 1994). Det kreves energi både til vandringen, dannelse av gonader og kjønnsdrakt. Og ikke minst er oppholdet i elven uten mat, og med kamp mot andre laks om de beste gyte plassene energikrevende (Fleming 1996).

Laksehanner kjemper om hunnene, ofte bare ved å markere sin størrelse (Fleming 1996). Ut fra dette burde både størrelse og fettreserver være viktige. Spesielt hunner bruker en høy andel energi på gytingen. På den annen side vil utsetting av kjønnsmodningen øke generasjonstiden, noe som vil gi laksen lavere «fitness».

En stor laks vil ha fordel av å lettere unngå predatorer fordi den svømmer raskere, og fordi den har færre fiender på grunn av sin størrelse. Dette kan tenkes å ha innflytelse på risikoprofilen for dødelighet i sjøen, og dermed påvirke reproduksjonsstrategien. Over tid har teoriene for hva som avgjør tidlig eller sen modning gradvis forandret seg. Sentrale begreper³ og teorier for denne hovedoppgaven er:

- Alder eller størrelse (Alm 1959).
- Genetikk (Nævdal *et al.* (1978a) og Gjerde (1986)).
- Fenotypisk plastisitet.
- Terskel (Thorpe 1994).
- Beslutningsvinduet (Thorpe 1989).
- Veksthastighet og ikke størrelse (Thorpe 1994).
- Teorier basert på livshistoriestrategi (Stearns 1992).
- To-stegs beslutningsprosess (Taranger 1993, s. 31).



Figur 2.5: Mer kompleks modell av forholdet mellom faktorene i eksperimentet. Pile-
ne angir årsaksretning. Det kunne også gått pil fra lys direkte til tertmodning, og fra
vekstfase1 til tertmodning. Her gjelder modellen innen kjønn.

Tanken bak figur 2.5 er å illustrere noen av begrepene i dette delkapittelet. «Vekstfase2» blir her beslutningsvinduet hvor en terskel på størrelse, fettnivå eller veksthastighet avgjør hvorvidt fisken modner som tert. At det ikke bare er genetikken som er avgjørende er en form for fenotypisk plastisitet (evne til å reagere på miljø eller tilstand). Hvordan fisken konkret reagerer på f.eks. at det er gode vekstvilkår blir en reaksjonsnorm (basert på livshistoriestrategiteorier, se neste avsnitt).

Det er videre naturlig å betrakte formeringen til laksen som det ultimate mål i livet. Vekst må derfor sees som del av denne prosessen, og laksen er derfor på vei mot reproduksjon helt fra den starter livet. Oppstart av kjønnsmodning kan oppfattes som et valg mellom somatisk vekst og investering i reproduksjon. Her kommer teorien om livshistoriestrategier til sin rett ved at den kan forutsi en optimal strategi, gitt rette premisser.

2.1.11 Livshistoriestrategier

Livshistoriestrategier vil si at man ser på hva som er lønnsomme måter å «gjennomføre livet på». Ofte blir det brukt matematiske modeller. Innenfor modellenes rammer vil man optimalisere laksens «fitness». Måtene å løse livet på som veies mot hverandre kalles «trade offs». For mer utdypning av tema se Stearns (1992) og Roff (1992).

Alder ved smoltifisering kan også oppfattes som en «trade-off». Det samme gjelder valget mellom å bli en dverghann eller en smolt. Eventuelt kan en dverghannen satse på samme strategi neste sesong, eller vandre ut i havet som smolt påfølgende vår eller senere.

Avveiiinger («trade-offs») med hensyn på reproduksjon kan være:

³Referansene må ikke sees som opphavsmenn til begrepene, men det er referanser der disse «begrepene» brukes.

- Alder ved kjønnsmodning.
- Tidlig eller sein gyting innen sesongen.
- Stor eller liten investering i reproduksjon.
- Mange små egg eller få store.
- Reproduksjon spredt på flere sesonger eller «alt» på engang.
- Antall gytegroper.

En organismes evne til å reagere på forandringer i miljøet kalles fenotypisk plastisitet. Hvordan organismen reagerer styres av reaksjonsnormer. Det vil si arveegenskaper som styrer hvordan en organisme reagerer, f. eks. på gode eller dårlige vekstvilkår.

Videre vil organismen være begrenset av sin anatomi, sin fysiologi og sin evolusjonære for-historie. Slike begrensinger kalles «constraints».

Organismen kan videre være tilpasset et stabilt eller variabelt miljø. Strategiens suksess vil også være tetthetsavhengig eller avhengig av hvor mange individer som velger tilsvarende strategi. Tettheten av anadrome hanner og dverghanner vil for eksempel virke inn på deres respektive gytesuksess.

Et annet begrep som brukes er den evolusjonære stabile strategi (ESS). Man tenker da at det finnes en optimal strategi, f. eks. vedrørende «alder» ved kjønnsmodning. En evolusjonær stabil strategi tenkes å være optimal slik at den ikke «invaderes» av andre strategier (gitt et sett av forhold som kan påvirke strategiene). Hos laksen eksisterer anadrome laks og dverghanner side om side i et område av en elv, gjerne fra samme opphav. Dette kan også være en ESS, men basert på forholdet mellom antallet anadrome laks og dverghanner. Dette har latt seg forklare best ved bruk spillteori («game theory»). Det er ikke bare én strategi som er lønnsom, men flere i et balansert forhold.

Dette vil si at det finnes modeller eller teorier som kan hjelpe oss til å forstå hvorfor laksen smoltifiserer og gyter ved svært ulik alder. Modellen vil predikere noen strategier, gitt visse betingelser. Vi kan så undersøke om prediksjonene stemmer i praksis. Er modellene riktig spesifisert kan de også brukes til å si noe om hva som er «meningsfulle» hypoteser.

2.2 Litt om biologisk tolking

I forsøket er det brukt en norsk oppdrettsstamme. Resultatene kan også være relevante for villlaks. Vi kan også se på laksens økologi for å forstå oppdrettslaksen. Laksen i Norge hadde i 1987 bare vært ca. 3–4 generasjoner i oppdrett, og det vil da være rimelig å anta at den ikke er så ulik villlaksen.

Jeg forsøker å skrive vill- eller oppdrettslaks, da jeg tror vi helt klart må skille mellom disse bestandene. Laksestammer kan ha svært ulike tilpasninger, og derfor er resultater fra én stamme ofte av begrenset gyldighet. Likevel vil jeg trekke inn resultater fra andre arter, spesielt laksefamilien.

Et annet forhold som er viktig vedrørende tolking av resultatene er forskjellen i miljøet til oppdrettsfisk og villfisk. Det er forskjell på et kar og en elv (eller et vann), og det er forskjell på en merd (kar i vårt datasett) og havet.

I karet er parren presset tett sammen slik at de ikke i samme grad som i elven kan være territorialer. I merdene er også laksen presset sammen på ett relativt lite volum, og kan bare svømme i merden. Det medfører at de ikke kan vandre mot elven når modningen starter.

Viktigst er kanskje at de får ferdig fôr i form av pellet i oppdrett. Generelt vil dette også ha konsekvenser for i hvilken grad resultatene er overførbare til villfisk i naturen. Vi kan oppsummere med at oppdrettslaksen har vesentlig forskjellige forhold enn laks i naturen m.h.p. mattilgang (pellets) , tetthet (sammenpresset) og manglende bevegelsesfrihet ⁴. Muligens er det lettere å bruke kunnskap fra laks i naturen til å forstå oppdrettslaks, enn det er å gå motsatt vei (avhengig av problemstillingen).

2.3 Biologiske problemstillinger

2.3.1 Lys

I denne oppgaven ser jeg på effekten av tre ulike lysbehandlinger. Dette gjøres i kapittel 8. Effekten av lys på vekstmønsteret er behandlet grundig av andre. Derfor er dette en underordnet problemstilling, men siden behandlingene påvirker fisken, må man ta hensyn til denne faktoren i de andre problemstillingene.

⁴Fisken kan ikke vandre fritt eller velge vannkvalitet.

I tillegg til at lysbehandlingene kan påvirke tidspunktet for smoltifiseringen, vil også veksten kunne påvirkes. Lysperioden antas forøvrig å kunne påvirke veksten gjennom endring i årsrytmen, eller som et direkte lysstimulus ved økt daglengde.

2.3.2 Kjønn

Jeg vil undersøke om hanner og hunner har ulikt vekstmønster. Dette gjøres i kapittel 8. men også i kapittel 9. Den grafiske framstillingen i kapittel 6. viser vesentlige resultater.

Generelt har ofte hanner en annen livshistoriestrategi enn hunner. Hos anadrom laks er det som regel hanner som blir modne i ferskvann (dverghanner). Postsmoltmodne er også som regel hanner. Andelen laks som modner som tert rapporteres generelt å være høyere hos hannene enn hos hunnene. Det er derfor ikke unaturlig å forvente ulikt vekstmønster hos hanner og hunner.

2.3.3 Modning

Veksten antas å påvirke tidspunktet for kjønnsmodning. Likeledes er det kjent at kjønnsmodningen påvirker veksten. Det er derfor naturlig å se nærmere på sammenhengen mellom vekst og kjønnsmodning. Dette gjøres i kapittel 10.

Vekstforskjeller mellom moden (tert) og umodenfisk (mellomlaks eller storlaks) er den biologiske hovedproblemstillingen i min tilnærming til datasettet. De to siste målingene er ikke interessante i forhold til å forklare hva som medfører tidlig modning (fordi modingen da er godt igang). Jeg så det også nyttig i den forbindelse å snevre inn antall variabler som analyseres, slik at eventuelle «effekter» ikke druknet i «ikke effekter» (omtalt i kap.7.). Derfor utelukkes de to siste målingene fra alle statistiske analyser, men jeg viser noe grafikk fra disse målingene.

2.3.4 Klassifisering

Laksen klassifiseres i hanner og hunner, i dverghanner og postsmoltmodne, samt tert og MSW-laks. Hvordan dette er gjort blir dermed en del av fundamentet for hovedproblemstillingen med vekstforskjeller mellom moden og umodenfisk.

Fiskene ble klassifisert som modnende (tert) ut fra gonadevekten og GSI (Gonado Soma-tisk Indeks). GSI er definert som gonadevekten delt på våtvekten til fisken. Våtvekt er vekten av fisken (våt fisk) før bløgging og sløyning, altså med blod og innvoller.

Det kan være usikkerhet knyttet metodene. Spesielt gjelder dette klassifiseringen i mod-nende på grunnlag av gonadevekten og GSI. Derfor behandles dette i et eget kapittel (nummer 4.)

2.3.5 Vekstkurver

Laksen har en relativt kompleks vekstkurve som er bestemt av ytre og indre faktorer. En vekstkurve vil derfor nødvendigvis inneholde mange parametre for å være tilnærmel-sesvis realistisk. Jeg vil se rent teoretisk på ulike vekstkurver. Laksens vekstmønster kan få konsekvenser for måten man gjør en eventuell parametrisering av kovariansmatrisen. Dette gjøres i kapittel 5. og 7.

2.3.6 Ny måleteknologi

Bakgrunn

Automatisk måling av relativt stor fisk i vann, er i dag en kjent teknologi. «Rammer» som «fotograferer» fisken fra flere vinkler når fisken svømmer gjennom, brukes av kommersi-elle aktører. De bruker rammene til å få ut størrelsesfordelingen på fisk i merd, til videre bruk i slakteplaner, produksjonsplanlegging og logistikk.

Problemet er at den kjente teknologien ikke kan identifisere enkeltindivider. Da kan man ikke få gjentatte målinger på samme individ, eller f.eks. identifisere avstamming eller «behandling» i forsøk. Dette gjør den kjente teknologien lite egnet (uegnet) til forsk-ningsformål. I forsøk håver man fisken opp av vannet, bedøver fisken før man veier og måler. Dette er arbeidskrevende, stresser fisken og medfører derfor vanligvis at man tar få målinger. Dette går igjen utover kvaliteten til forskningen. Stressingen av fisken ved gjentatte målinger påvirker antageligvis også resultatene.

«Teknologien»

Det er mulig å bruke måleteknologien som «fotograferer» fisken fra flere vinkler når fisken svømmer gjennom. Kobler vi denne teknologien til lesbare individuelle merker får vi målt fisken uten håndtering. Man får da et lengde og vekt estimat av individuelle identifiserbare fisk.

Det finnes flere typer merker, både indre og ytre. Med et ytre merke menes et merke festet til fisken ytre og med indre merke menes et merke anbrakt inne i fisken. F. eks. kan merket være en Passiv Integrated Transponder (dvs. PIT-merke). PIT-merkene ⁵ legges i buken av fisken, og har en vid bruk i dag.

Dersom vi ikke ønsker å tvinge fisken gjennom måleapparatet, slik at de går gjennom av seg selv, får vi ulike måletidspunkter for ulike fisk. Det får alvorlige konsekvenser for eventuell statistisk bearbeiding av dataene. Da kan vi individualisere parametriseringen av kovariansmatrisen, slik det er beskrevet i Diggle (1990). Jeg har også foreslått måter å forbedre parametriseringen av kovariansmatrisen slik at dette er mer tilpasset biologien til laksen (se avsnitt 7.6.6).

⁵PIT-merke er blitt brukt i det forsøk som behandles i denne oppgave.

3 Om forsøket

Forsøket er i hovedsak beskrevet i følgende publikasjoner:

- Stefansson *et al.* (1991)
- Vestfossen *et al.* (1991, s. 292)

Jeg vil derfor i dette kapittelet bare gi en kort beskrivelse av forsøksoppsettet og utfyllende opplysninger.

3.1 Forsøksfisk

I forsøket er det brukt Atlantisk laks (*Salmo salar*) fra Havbruksstasjonen Matre. Familieopphavet er blandet, og det er ikke registrert hvilken familie den enkelte fisk kommer fra (Ragnar Nordtvet pers. med.). Stamfisken ble strøket i 1986, mellom siste uken i november og midt i desember. Etter klekking gikk fisken på kontinuerlig lys frem til forsøkstart i juli 1987.

Forsøksfisken er i prinsippet et tilfeldig utvalgt fra en del-populasjon av all fisken ved Matre havbruksstasjon. En vanlig praksis i smoltoppdrett er at den minste parren er sortert vekk og destruert. Jeg har ikke kjennskap til om det er gjort i dette tilfelle.

Forsøksfiskene ble i juli sortert i 5 størrelsesgrupper ved hjelp av sorteringrister. De fiskene som går gjennom en rist med 7 mm lysåpning hører til størrelsesgruppen 7-, mens den som blir holdt igjen hører til 7+. Fiskene ble sortert i størrelsesgruppene 7-, 7+, 8+, 9+ og 10+. Fra hver av disse 5 størrelsesgruppene ble det håvet ut 180 tilfeldige parr til bruk i forsøket. Tilsammen 900 parr.

Ved forsøksstart ble det altså gjort et tilfeldig utvalg av parr fra en (del-)populasjon. Parren var ett utvalg fra ulike størrelses sorteringer, og representerer således ett spesielt utvalg. Dette spesielle utvalget medfører at vi får ett flatere utvalg av parren enn ved tilfeldig trekk fra hele populasjonen. Dette kan faktisk være gunstig. På den annen side

er ikke utvalget representativt for en vanlig situasjon i akvakultur, der parren er sortert i ulike størrelsesgrupper (i hvert kar er det da parr av ganske lik størrelse).

3.2 Merking

Fiskene fra hver størrelsesgruppe ble individuelt merket, altså totalt 900 individer.

Merkene ble levert fra Fish Eagle. De er små magnetiske brikker som inneholder en 12 sifferet alfanumerisk kode. Merkene ble satt inn i bukhulen på hver enkelt fisk under bedøvelse med benzokain. Selv om merkene er inni fisken lar de seg lese av utenpå. Merkene inneholder en «Passiv Integrated Transponder», og kalles derfor PIT-merker. Den alfanumerisk koden leses av ved bruk av en PIT-nummerleser.

3.3 Forsøksoppsett

I juli 1987 ble parren fordelt i 6 like kar for lysbehandling. I hvert kar ble det tilfeldig plukket ut 30 merkede fra hver størrelsesgruppe, slik at hvert kar inneholdt 150 merkede parr.

Lysbehandling ble umiddelbart satt i gang. To av karene fikk naturlig lysregime (SLN). To fikk vekslende naturlig lysregime pluss konstant tilleggsls (SLN+B) og de to siste karene fikk 24 timer konstant lys (LD24:0).

Lysbehandlingen pågikk frem til 5. mai 1988. Deretter ble fisken overført til et felles kar med naturlig lys og gradvis tilvennet økende saltholdighet i løpet av en uke. Her var fisken fram til avslutning av forsøket i september 1989.

3.4 «Lyskvalitet»

Lysintensitet er beskrevet i Stefansson *et al.* (1991) og der oppgis lyskildene å være beskrevet i Stefansson og Hansen (1989) .

3.5 Fôring

Det ble foret med kommersielt tørrfôr (Skretting Tess Elite, 3.0G) ved hjelpe av skiveautomater fra Akvaprodukter. Det ble kun fôret når det var lys. I ferskvannsfasen ble de karene med kortest daglengde, bestemmende for hvor lenge det skulle fôres hver dag. Dermed ble gruppene med simulert naturlig lys, begrensende for hvor lenge det ble fôret en bestemt dag. Det ble fôret etter en tabell (Austreng *et al.* 1987) som tar hensyn til temperatur og fiskestørrelse (justert en gang pr. uke).

3.6 Helse

Det er ikke registrert om fisken vært undersøkt av fiskehelsebiolog eller veterinær i forsøksperioden. Vi vet derfor ikke om noe av fisken kan ha hatt en form for sykdom eller produksjonlidelse. Dette kan f.eks. ha vært blakking av øyelinsen (som svekker synet), ryggdeformasjoner som gir avvikende fasong på fisken, costia-infeksjon som angriper gjellene eller annet som kan ha svekket veksten.

3.7 Veiing og måling

Fisken ble sultet en dag i forkant av veiing og måling. Vekten ble avrundet til nærmeste gram i ferskvannsfasen og nærmeste 10 gram i sjøvannsfasen. Dette har sammenheng med presisjonen til veieapparatene (vektene) som ble brukt var ulik i de to fasene.

I og med at fiskene er våte overestimeres vektene konsekvent. Det følger med mer vann jo større fiskene blir. Også i gjellene vil det være en del vann. Dette antar jeg ikke spiller så stor rolle for forholdet mellom fiskene. Man om man tolker vekt som ett absolutt mål, vil det ha større betydning.

Lengden ble målt fra snute til midt i sporen på et målebrett, og rundet ned til nærmeste hele cm. i sjøfasen. Under veiing og måling var fisken bedøvd med benzokain.

Ut fra tabell 3.1 på neste side og tabell 3.2 på neste side antar jeg at jeg har måling nr. 1, 5, 7, 8, 9, 10 og 11 hos Stefansson *et al.* (1991). Jeg velger å bruke datoene fra mitt eget utkast.

Dato for operasjon	Nr.	Dager fra forrige måling	Annet
15 september 1987	1	-	Fordeling i 6 kar
1 mai 1988	-	-	Ett felles kar
26 mai 1988	2	254	-
15 september 1988	3	112	-
15 desember 1988	4	91	-
8 mars 1989	5	83	-
9 mai 1989	6	62	-
24 juli 1989	7	76	-
27 september 1989	8	65	Utslakting, Veiing av gonader

Figur 3.1: Tabell av måletidspunkt: Ut fra Vestfossen (1991) og utkast til oppgave.

Dato for operasjon	Nr	Dager fra første måling	Annet
21 november 1987	1	0	Fordeling i 6 kar
23 desember 1987	2	-	-
28 januar 1988	3	-	-
3 mars 1988	4	-	-
26 april 1988	5	157	-
5-12 mai 1988	-	-	Saltholdighet økt
15 juni 1988	6	207	-
12 september 1988	7	296	-
12 desember 1988	8	388	-
15 mars 1989	9	481	-
15 juni 1989	10	572	-
12 oktober 1989	11	692	Utslakting, Veiing av gonader

Figur 3.2: Tabell av måletidspunkt: Ut fra Stefansson *et al.* (1991)

3.8 SGR og kondisjonsfaktor

For et kortere tidsrom vil en kunne anta at fisk vokser en fast prosent av sin kroppsvekt pr. dag. Antar vi at hvert enkelt individ vokser en fast prosent av sin kroppsvekt pr. dag i perioden $t_1 - t_0$ får vi ligningene:

$$W(t_1) = (t_0)e^{g(t_1-t_0)}$$

$$SGR = (e^g - 1)100, \quad \text{der } g = \ln W(t_1) - \ln W(t_0) / (t_1 - t_0)$$

der SGR = Specific Growth Rate = fast prosentvis økning av vekten pr. dag (døgn) i perioden. Dette er den SGR jeg henviser til, men det også vanlig å bruke SGR definert som g i ligningene overfor men modifisert ved at man ganger med 100. Vi får da ligningen:

$$\hat{SGR} = 100(\ln W(t_1) - \ln W(t_0)) / (t_1 - t_0)$$

Kondisjonsfaktor er definert som:

$$\text{Kondisjonsfaktor} = 100W/L^3$$

der W er vekt i gram og lengden er oppgitt i centimeter. Denne betegnes også som Fultons kondisjonsfaktor.

3.9 Klassifisering i kjønn og modningskategorier

Laksen klassifiseres i hanner og hunner, i dverghanner og postsmoltmodne, samt tert og MSW-laks. Dette blir omtalt i et eget kapittel 4, kalt klassifisering.

Ved forsøksavslutning ble gonadevekten registrert. Fiskene ble klassifisert som modnende (tert) ut fra gonadevekten og GSI (Gonado Somatisk Indeks). Det kan være usikkerhet knyttet metodene. Spesielt gjelder dette klassifiseringen i modnende på grunnlag av gonadevekten og GSI. (Det ble slaktet ett mindre antall fisk før forsøksavslutning også, men disse er ikke tatt med i materiale).

3.10 Dataregistrering

På hvert måletidspunkt er vekt, lengde og fire siffer fra den alfanumeriske koden registrert på vannfast papir. Man antok at fire av sifrene var en entydig identifikasjon av individet. Samtidig ble PIT-numrene overført via kabel inn på en datamaskin fra PIT-nummerleseren. Deretter ble registreringene lagt inn i statistikkprogrammet RS/1. Omtrent halvparten var lagt inn før jeg kom inn i bildet og resten er gjort av meg.

Grunnet feilregistreringer av de fire sifrene ved nedskrivningen på papir, får man en del observasjoner med identifikasjonsnummer som ikke stemmer. Disse observasjonene kjenner man da ikke identiteten til. Disse observasjoner kunne ikke brukes mhp. problemstillingene, og gikk altså ut av materialet på grunn av ukjent identitet.

3.11 Kvalitetsikring av dataene

Utfra spredningsplot av dataene mellom måletidspunktene, kunne jeg se at det var svært mye «støy» mellom tre etterfølgende måletidspunkt. Jeg forstod at noe måtte være feil.

I datalistene står hver fisk på en rekke med måleserien fortløpende. Der fantes en del fisk som enten var svært små eller svært store. Ut fra datalistene var det mulig å se at på det aktuelle måletidspunktet hadde en del av målingene forskjøvet seg vertikalt.

Ved å bruke svært små eller svært store fisk som riktig referanse, kunne jeg flytte målingene vertikalt. Dette må gjøres med riktig sortering, og at man går tilbake tilbake til «råfilene».

Måleserien til referansefiskene ble da liggende på en riktig rekke. Spredningsplot og SGR inneholdt da ikke mer støy enn målingene ellers. Referansefiskene gjorde det mulig å bruke et måletidspunkt mer fra materialet enn ellers.

3.12 «Rensking» av uteliggere

Kvalitetetsikringen er ellers gjort hovedsaklig ut fra biologiske kriterier. Jeg luket ut observasjoner som er biologisk umulig eller svært usannsynlige. Dette ble gjort før de statistiske analysene.

Først har jeg brukt fordeling av lengde og vekt til å lokalisere «uteliggere» . Deretter brukte jeg kondisjonsfaktor til å se på forholdet mellom vekt og lengde, slik at jeg kunne lokalisere urimelig «feite» eller «tynne» fisker. Jeg gikk da tilbake for å finne eventuelle feilpunchinger på enten lengde eller vekt.

Jeg valgte å bruke SGR til å finne fisk som vokste usedvanlig mye eller lite, både med hensyn på vekt og lengde. Fisk med SGR under $-0,3$ prosent m.h.p. vekt ble utelatt og fisk med negativ lengdevekst. Like så godtok jeg fisk med kondisjonsfaktor mellom 0,6 og 1,8. Grensene for kondisjonsfaktor ble dengang regnet som vide, men jeg kunne bare konstatere at fisker med ekstreme kondisjonsfaktor hadde dette på «alle» måletidspunktene.

Videre brukte jeg spredningsplot med «konfidens-ellipser» til ytterligere finne «uteliggere». En lignende prosedyre anbefales (Johnson og Wichern 2002, s. 190) , men at man også ser på generalisert kvadratisk distanse: $(Y_i - \bar{Y})^T S^{-1} (Y_i - \bar{Y})$. Deres prosedyre ser ut til å skulle brukes på rådataene og er basert først og fremst på statistikk. Jeg antar at samme prosedyre er godt egnet til å bruke også på residualene fra modellen.

Å sette grenser for hva som er en urimelig observasjon er ikke lett. Noen holdepunkter har vi. En fisk vil ikke krympe i lengde selv om den sultes. En fisk vil også vokse i lengde selv når den taper vekt. Å sette en øvre og nedre grense for kondisjonsfaktor er verre. Likeså en øvre grense for vektøkning. Øvre grense for lengdevekst er lettere, fordi dette er en jevnere prosess.

Dessverre har jeg ikke noe fullgod dokumentasjon av dette arbeidet, men det kan eventuelt dokumenteres ved å sammenligne «råfiler» med den jeg nå bruker. Det ligger et betydelig arbeid bak å få datasettet brukbart.

4 Klassifisering

I dette kapitlet ser jeg på hvordan laksen klassifiseres i hanner og hunner, i dverghanner og postsmoltmodne, samt tert og MSW-laks.

4.1 Klassifisering i hanner og hunner

Klassifisering som hanner og hunner er gjort på grunnlag av visuell inspeksjon av gonadene. Gonadene er de indre organer, der henholdsvis melke og rogn blir produsert. Hannenes gonader er gråhvite i front, mens hunnenes gonader er blekt oransje og med anlegg til små rognkorn.

Metoden er svært sikker. Det var kun tre fisk som hadde så lite utviklet gonader at vi ikke kunne bestemme kjønn.

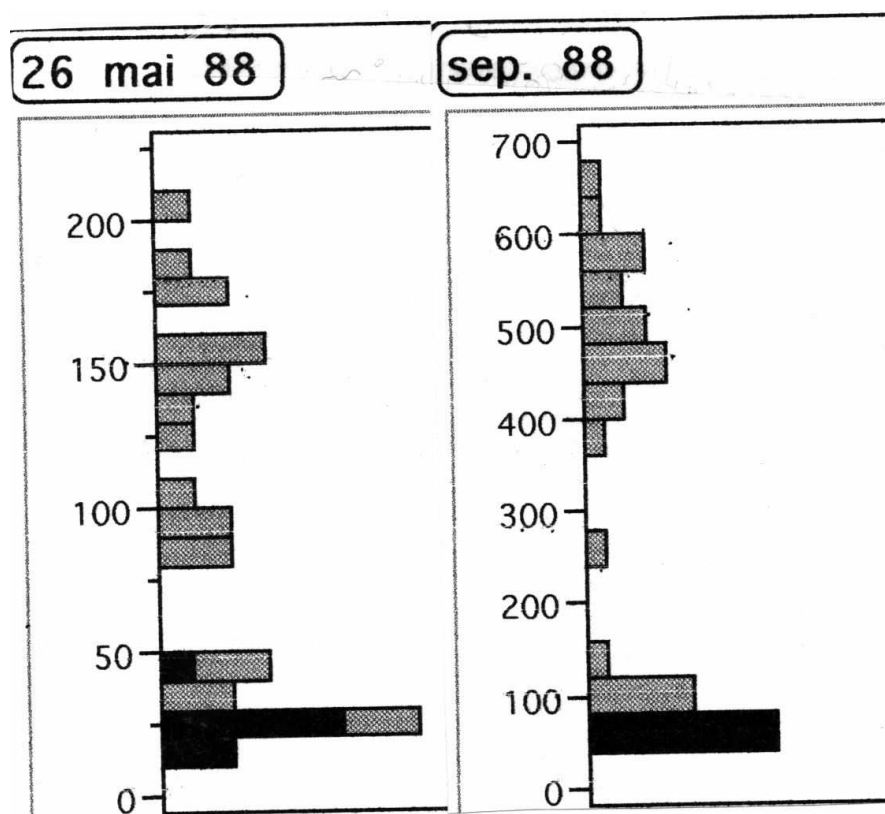
4.2 Klassifisering som postsmoltmodne

«Postsmoltmodne» har blitt klassifisert ut fra ytre kjennetegn (farge og fasong). Figur 4.1 på neste side viser at det ser ut til å være to størrelsesfordelinger. Det er mulig at de minst egentlig er dverghanner.

4.3 Klassifisering som tert

Fiskene ble klassifisert som modnende (tert) ut fra gonadevekten og GSI (Gonado Soma-tisk Indeks). GSI er definert som gonadevekten delt på våtvekten til fisken. Våtvekt er vekten av fisken (våt fisk) før bløgging og sløying, altså med blod og innvoller. Listen nedenfor gir forklaring av flere begreper, og forkortelser brukt i tabell 4.3.3.

- MSW = «Multi Sea Winter»



Figur 4.1: Histogram av vekt hos fisk som er klassifisert som modne første høsten i sjøen. De er klassifisert ut fra ytre kjennetegn.

- Vi har tert (1SW), mellomlaks (2SW) , storlaks (3SW etc.). Tallene i kodene angir er antall vintre i sjøen før modning. To eller flere vintre er MSW.
- PSM = postsmoltmodne
 - Gytemoden etter en sommer i sjø.
- MM = multimodne (NB! ikke etablert begrep)
 - Modne både som postsmolt (eventuelt dverghann) og tert.

Vedrørende modning ser jeg videre først på hvordan dette er gjort generelt og i andre utvalgte arbeider. Deretter ser jeg hvordan dette ble gjort i mitt datasett, og så tilslutt ser på alternative grenser i mitt datasett. Se forord (1.7) for begrunnelse for at opprinnelig inndeling i modne og umodne beholdes.

Målsetningen er også å få frem at det kan være usikkerhet knyttet metodene. Spesielt gjelder dette klassifiseringen som modne på grunnlag av gonadevekten og GSI. Klassifiseringen er sikrere når flere metoder kombineres.

4.3.1 Klassifiseringmetoder

Laksen kan klassifiseres som moden eller umoden ut i fra ytre kjennetegn, hormonnivå, ved visuell vurdering av gonadene (eventuelt mikroskopi/histologi) eller ut i fra en gonadosomatisk indeks (GSI). Det beste er om mulig å kombinere metodene overfor. Vurderingene gjøres som regel på grunnlag av plott og histogrammer (se figur 4.4 på side 45 og figur 4.5 på side 46).

Det er utbredt å bruke en bimodal fordeling av GSI som grunnlag for å klassifisere. Generelt har man da om sommeren eller høsten klassifisert laks med små gonader eller liten GSI som umoden.

Det kunne være mulig å bruke clustring, diskriminantanalyse eller andre metoder (se f.eks. Ramsay og Silverman (1997)) til hjelp eller for å gjøre dette mer formelt. Dette blir da en støtte for de biologiske vurderinger.

Diskriminantanalyse, clustring, etc. har jeg ikke sett meg tjent med å bruke. Et ordtak sier «Du skal ikke skyte spurv med kanoner». Jeg gjør en grundig grafisk presentasjon, og anser det som god nok for denne oppgaven.

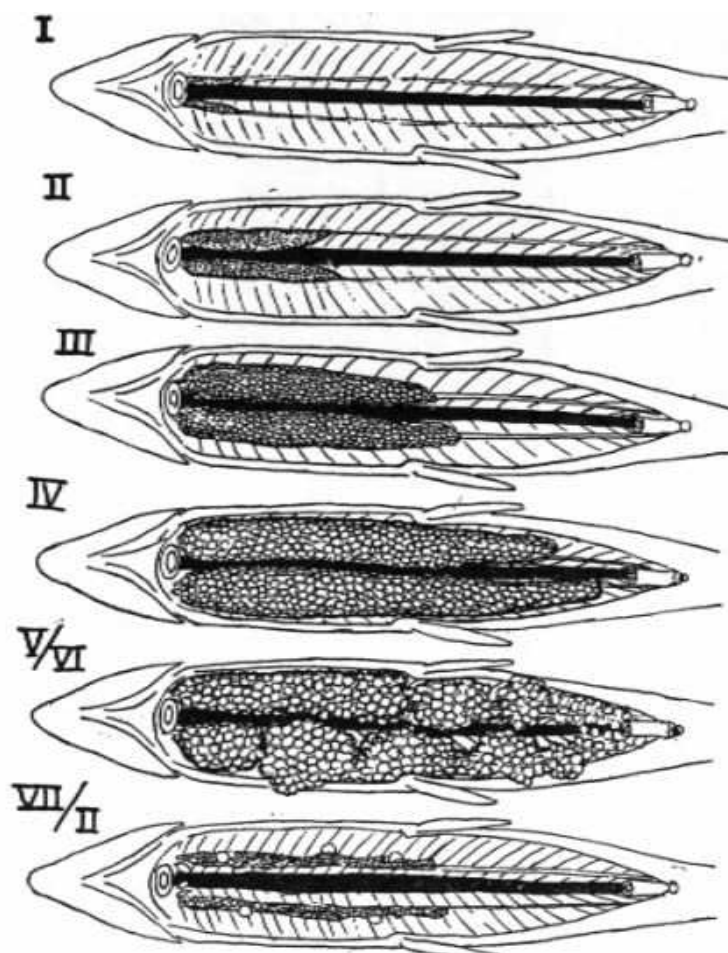


Fig. 118. Forskjellige stadier i utviklingen av rogonna hos hunnørret fra tidlig stadium til utgytt fisk. Romertallene angir betegnelse på stadiene (se teksten).

Figur 4.2: Illustrasjon gonader hos ørret (hunner). Hos laks er gonadene ikke vesentlig forskjellig. Nummeringen har med utvikling utover i gytesesongen (1, 2 og 3 er mest illustrerende for tidlige stadier innen «gytesesongen»). Modifisert etter Sømme (1954).

4.3.2 Grenser i utvalgte arbeider

Thorpe (1994) har i figur 4.3 på neste side vist utviklingen av GSI hos hunner og hanner. I november/desember, omtrent et år før gytingen, kan det se ut som at «terten» får en brå økning av GSI.

Grensene mellom moden og umoden settes ulikt hos ulike forfattere. I tabellene nedenfor har jeg angitt grensene som er brukt i noen utvalgte publikasjoner.

Som vi ser av tabellene settes grensen lavere hos hannene enn hunnene, fordi gonadene til hannene er mindre enn hos hunnene. Det kan også se ut som grensene settes høyere desto senere på året laksen er slaktet. Når GSI er vurdert og ikke gonadevekten, må grunnen være at gonadene til den laksen som ikke skal bli moden inneværende sesong, også vokser proporsjonalt mer enn resten av fisken.

Forfattere	Årstid	Alder	Vekt fisk	GSI	Gonadevekt	Sted
Thorpe <i>et al.</i> (1990)	1 juni	1SW	ca. 0,8	0,1	–	Skottland
Kadri <i>et al.</i> (1996)	19 juni	1SW	–	0,10	–	Skottland
Oppedal (2002)	24 juni	1SW	ca. 4,0	0,18	10 gram	Matre
Kadri <i>et al.</i> (1996)	26 juli	1SW	ca. 2,7	0,20	–	Skottland
Endal <i>et al.</i> (2000)	september	1SW	0,9	0,4	3 gram	Matre

Tabell 4.1: Grenser mellom modne og umodne hanner i utvalgte arbeider. «Vekt fisk» er gjennomsnittsvæker oppgitt i publikasjonene eller estimert fra «figurer».

4.3.3 Grenser i vårt datasett

Ved terminering av forsøket ble laksen klassifisert som modne eller ikke ut i fra en kombinasjon av gonadevekten og GSI. (Gonadevekten ble da registrert.)

Forfattere	Årstid	Alder	Vekt fisk	GSI	Gonadevekt	Sted
Kadri <i>et al.</i> (1996)	19 juni	1SW	–	0,38	–	Skottland
Oppedal (2002)	24 juni	1SW	ca. 4,0	0,30	14 gram	Matre
Kadri <i>et al.</i> (1996)	26 juli	1SW	ca. 2,7	0,42	–	Skottland
Endal <i>et al.</i> (2000)	september	1SW	0,9	0,8	–	Matre

Tabell 4.2: Grenser mellom modne og umodne hunner i utvalgte arbeider. «Vekt fisk» er gjennomsnittsvæker oppgitt i publikasjonene eller estimert fra «figurer».

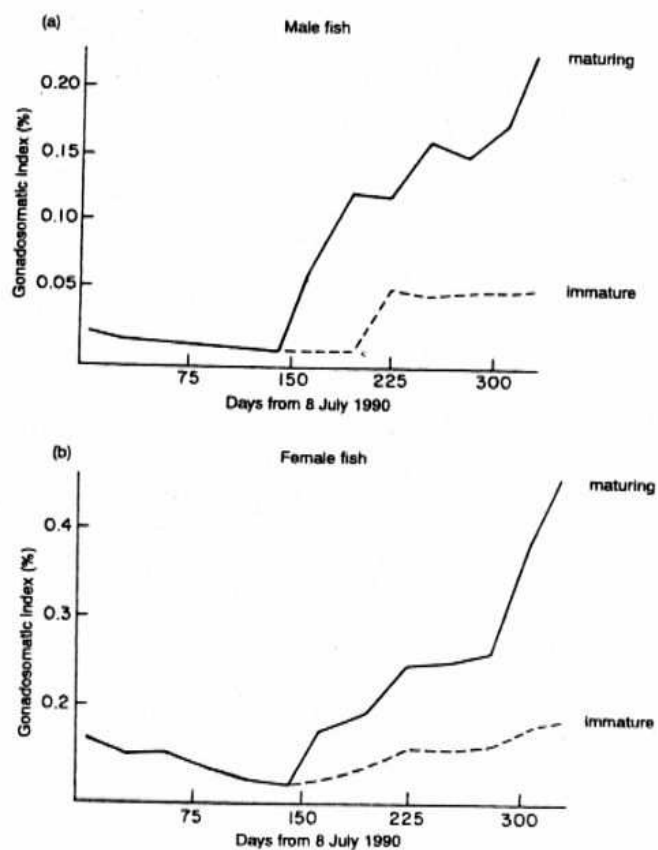
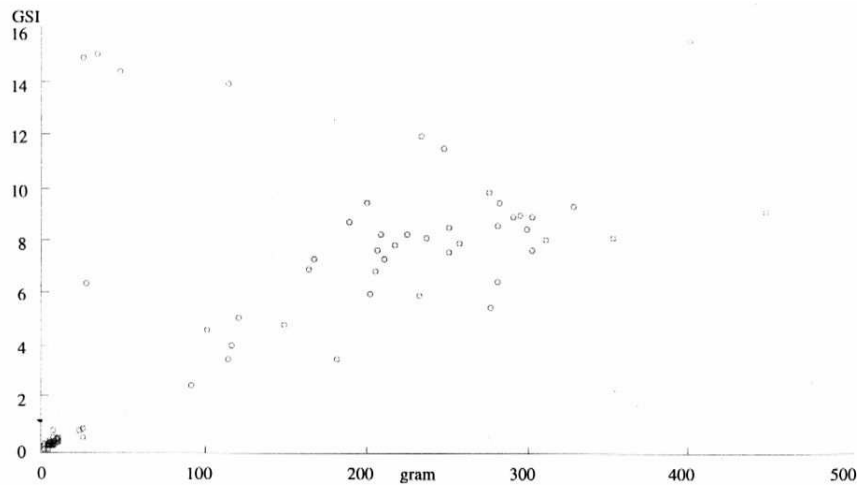


Figure 2. Gonadosomatic index of maturing and immature salmon in sea cages: aged 1+ in July 1990; (a) male fish; (b) female fish.

Figur 4.3: Illustrasjon av GSI hos hanner (øverst) og hunner (nederst). Horisontal akse er antall dager etter 8 juli 1990. Jeg antar kurvene er fra utvalg av ulike fisk, og at «knekkpunktene» er noen av de reelle måletidspunkt. Modifisert etter Thorpe (1994).



Figur 4.4: Illustrasjon av gonadevekt mot GSI hos hanner, inkludert postsmoltmodne (i forsøket som behandles i denne oppgave).

Ved start av hovedfaget i akvakultur ble grensene satt etter råd fra Geir Lasse Taranger. Vi klassifiserte laks med små gonader eller liten GSI som umoden (mellomlaks eller storlaks) utfra en bimodal fordeling. Vi brukte spredningsplott av gonadevekten mot GSI.

- GSI=0,8 prosent hos hannene
- GSI=1,2 prosent hos hunnene

Mot slutten av hovedfaget i akvakultur forandret jeg disse. Grensene ble da satt til:

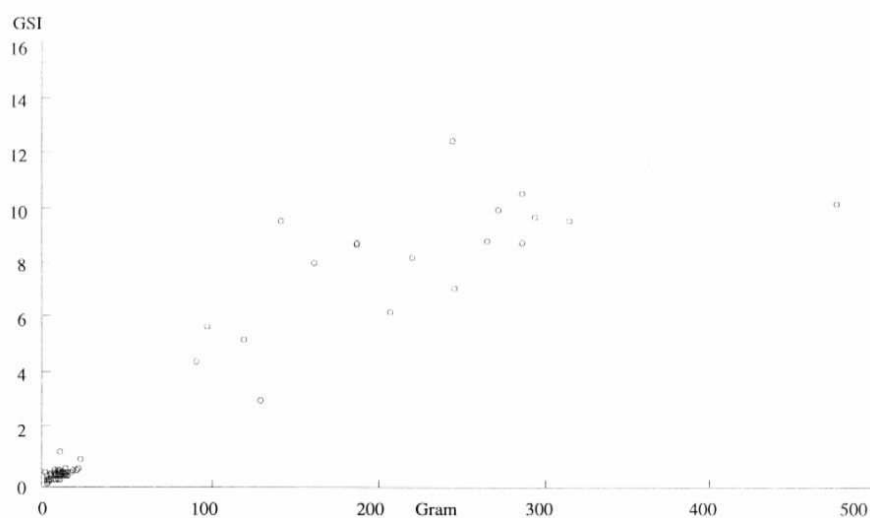
- GSI=1 prosent hos hannene
- GSI=1 prosent hos hunnene

Disse grensene ble kombinert med en grense på 10 gram i gonadevekt. Jeg klassifiserte da også noen fisker som usikre med hensyn på om de var tert eller ei, men jeg er usikker på hvor mange dette var. Uansett dreier det seg om noen få fisker.

4.3.4 Mulige alternative grenser i «vårt» forsøk

Hanner

På spredningsplottet hos hannene figur 4.4 kan det se ut til å være et skille i GSI mellom ca. 0,8 og 2,4. Dette fordi vi ikke har observasjoner i det området. På histogrammet



Figur 4.5: Illustrasjon av gonadevekt mot GSI hos hunner, inkludert en postsmoltmoden (i forsøket som behandles i denne oppgave).

Lysregime	Postsmoltm.	MM	Tert	MSW	Ukjent(tert/MSW)	Total
SNL	(14)	(1)	15	157	128	300
SNL+B	(7)	(2)	26	157	117	300
LD24:0	(12)	(1)	20	160	119	299
	33	4	61	474	364	899

Tabell 4.3: Antall postsmoltmodne (PSM), tert, multimodne (PSM og tert) og MSW. «Ukjent» er fisk som er gått ut av forsøket for terminering (unntatt «postsmoltmodne») Hanner og hunner slått sammen (fra tidligere utkast til oppgave i akvakultur). Rekkesommene fremkommer ved å ikke ta med tallene i parentes (NB! en fisk mangler). Forklaring av forkortelser er i første avsnitt i dette kapittel.

Lysregime	MSW:hanner	Tert:hanner	MSW:hunner	Tert:hunner	Ukjent	Total
SNL	64	12	93	3	128	300
SNL+B	62	19	95	7	117	300
LD24:0	67	13	93	7	119	299
	193	44	281	17	364	899

Tabell 4.4: Antall tert og MSW. Hanner og hunner hver for seg.

i figur 4.7 på side 49 øverst til høyre, over gonadevekten til hannene ser vi at 4 gram eller 10 gram kan være aktuelle grenser. Ut i fra figur 4.6 på neste side, kan en grense på 0,8 i GSI kombinert med 10 gram i vekt, være fornuftig.

Hunner

Hos hunnene 4.5 har vi ett tilsvarende område uten observasjoner mellom ca 1,2 og 2,9.

Dersom vi ser på histogrammet over gonadevekten til hunnene 4.7 kan fordelingen se ut til å være en grense på 14-15 gram (grovt anslått). Histogrammet av GSI 4.7 i midten antyder at 0,6 kan være en mulig grense.

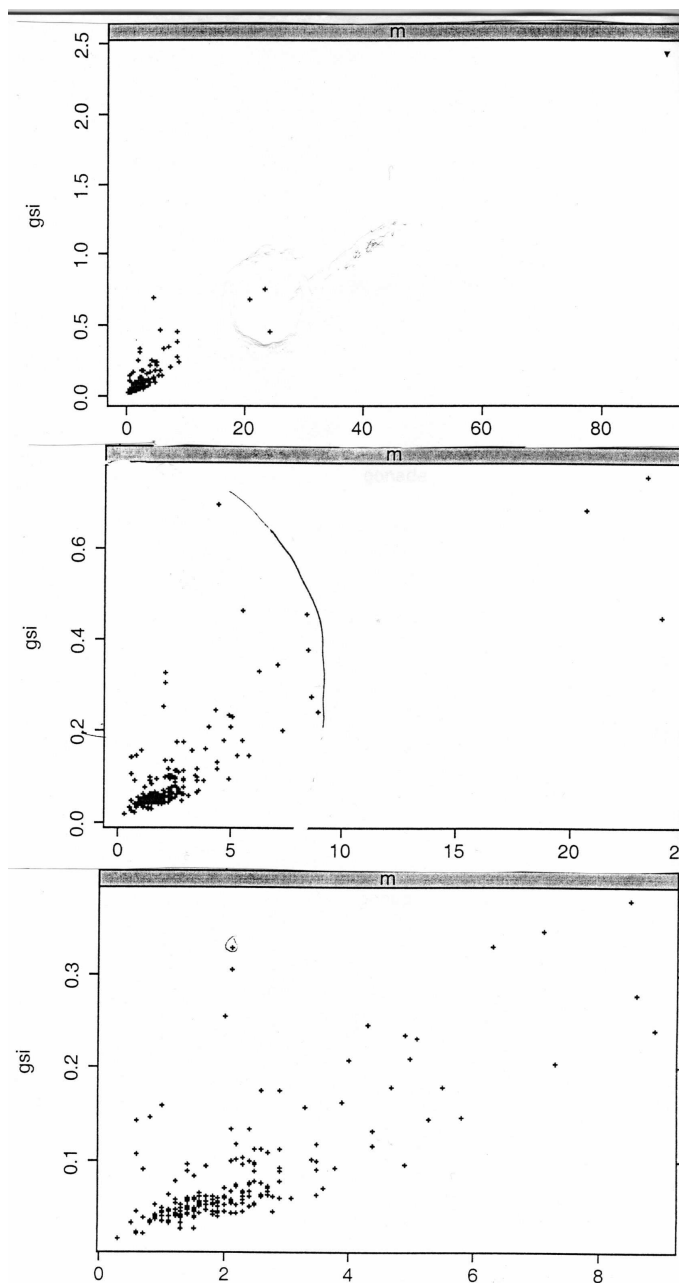
4.3.5 Diskusjon

Generelt er metoden forbundet med en viss usikkerhet. Jeg har gjort klassifiseringen ut i fra en vurdering som beskrives overfor. Vurderingen er gjort på et biologisk grunnlag ut i fra plott og histogrammer. Metoden lar seg videre vanskelig etterprøve, fordi laksen slaktes for å veie gonadene. Vi vet altså ikke absolutt sikkert om laksen modner inneværende sesong, selv om gonadene er «større». Usikkerheten kan også gjelde fisker i «grenseland» (som f.eks kan være «sene i sesongen»).

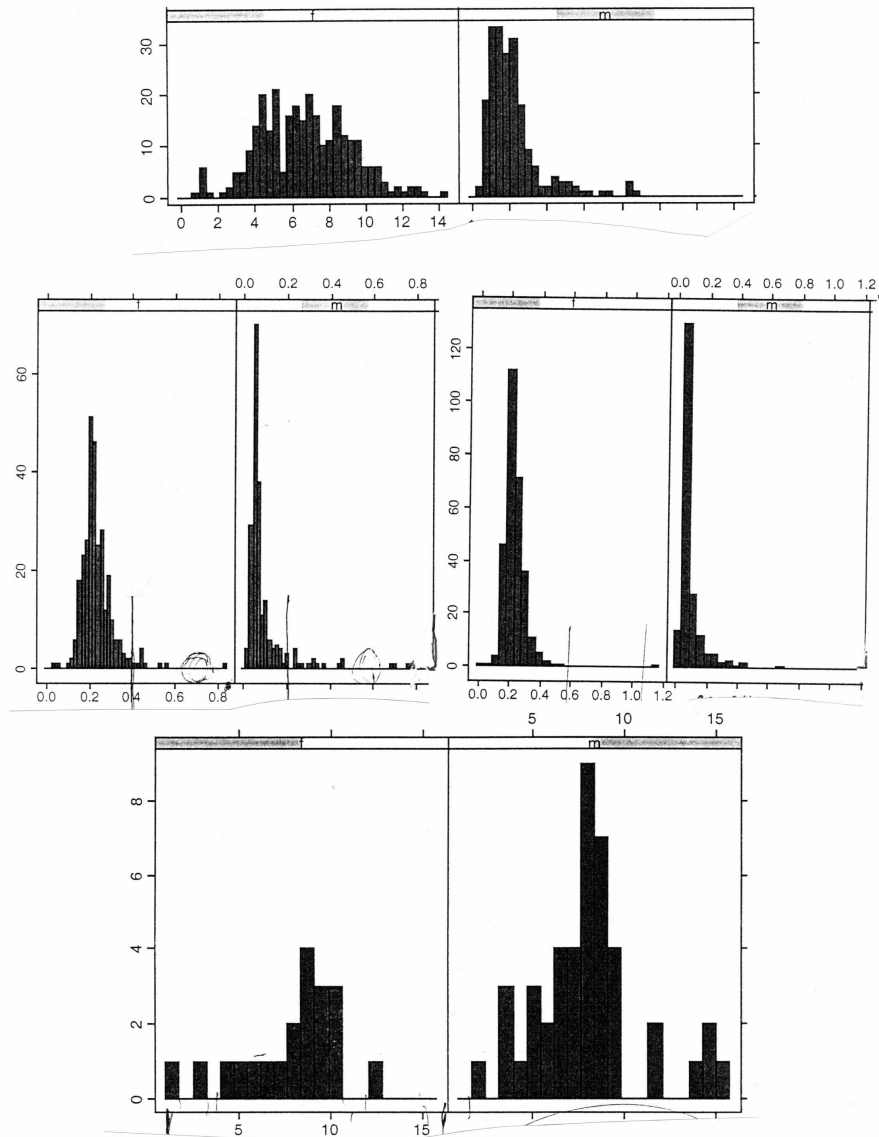
I figur 4.8 på side 50 ser vi at GSI og gonadevekten ligger i «etasjer». Dette er delvis knyttet til ulike gonadvikt hos hanner og hunner. GSI hos umodne hunner er større enn hos umodne hanner. I figur 4.9 på side 50 ser vi det er stor spredning i GSI hos fisk som er klassifisert som umodne. Det ser ut til være en viss utvikling av gonadene spesielt hos en del hanner.

Det blir også spekulert i om vi i september (året før gyting) egentlig har tre fordelinger av gonadevekten og GSI. Den tredje fordelingen kan da representere fisk som skal modne neste høst, men som allerede har startet gonadeutviklingen.

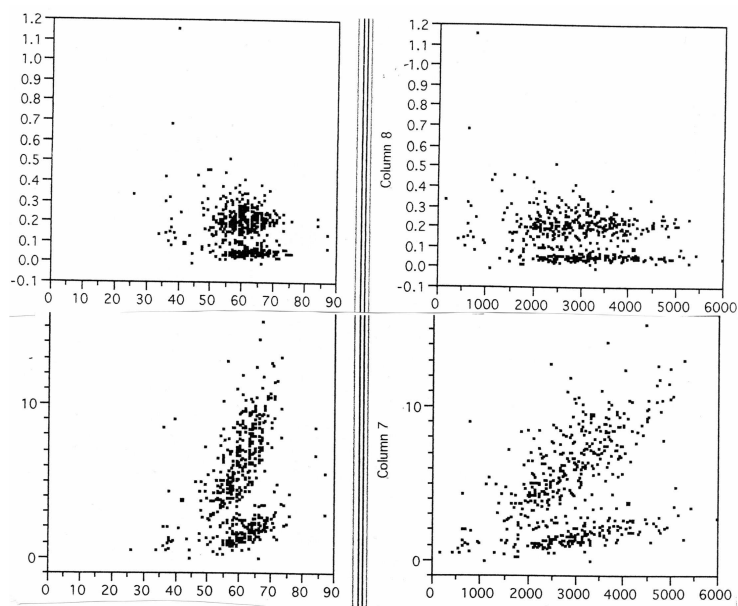
Dette støttes av Geir Lasse Taranger (pers. med.) som oppgir å ha opplysninger som støtter at en del hunner begynner modningen allerede tidlig september, mer enn et år før de skal gyte. Dette gjelder spesielt for mellomlaks (2SW). Det spekuleres videre også i om dette kan være et fenomen knyttet til oppdrett: at villfisk ikke ville ha startet modningen så tidlig (Gunner Nævdal pers. med.).



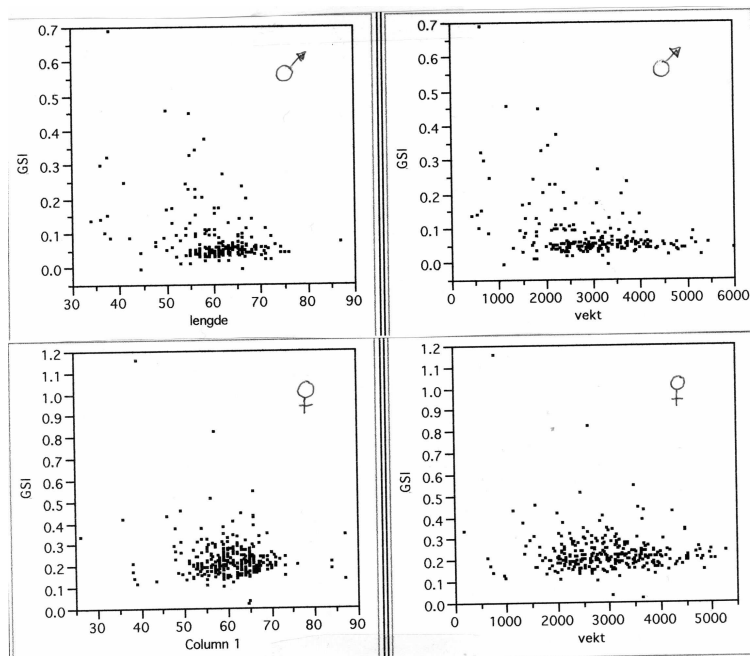
Figur 4.6: Illustrasjon av gonadevekt i gram (x-akse) mot GSI (y-akse) hos hanner, inkludert postsmoltmodne. Her har jeg «zoomet» gradvis mer inn i nederste hjørne til venstre. Linjen i midterste plott representerer en mulig grense.



Figur 4.7: Histogrammene er delt i hunner til venstre (f) og hanner til høyre (m): Øverst gonadevekt mot GSI (< 15 gram). Midten venstre GSI (< 1) og midten høyre GSI (< 1,2). Nederst GSI (> 1).



Figur 4.8: Illustrasjon lengde (venstre) og vekt (høyre) mot GSI (<1,2) øverst og mot gonadevekt (<15 g) nederst. Begge kjønn er inkludert.



Figur 4.9: Illustrasjon av lengde (venstre) og vekt (høyre) mot GSI. Hanner øverst og hunner venstre.

5 Vekstkurver

Poenget med dette kapitlet er å se på hvordan biologien til laksen påvirker vekstkurver, og å peke på muligheten til å bruke nlme-rutinen i Splus («Nonlinear mixed effects») til å inkludere individeffekter.

En utbredt innfallsvinkel for å beskrive vekst hos fisk er vekstkurver (Kolding og Bergstad 1988). Med vekstkurver menes en funksjon, og den er vanligvis ikke-lineær. Man antar da at vekstforløpet for gjennomsnittet kan beskrives ved hjelp av en formel. Eventuelt kan kurven tilpasses grupper av individer eller hvert enkelt individ.

I kommersielt oppdrett brukes vekstkurver i dataprogrammer for å planlegge, budsjetere og optimalisere driften. Fôringen styres som regel av en datamaskin, og da brukes også vekstkurver. Alternativt bruker man veksttabeller (Austreng *et al.* 1987). Ofte vil man da i praksis være interessert i gjennomsnittlig vekt og en fordeling rundt denne for hvert enkelt kar eller for hver enkel merd.

Jeg ser først på vekstkurver generelt før jeg diskuterer den såkalte powermodellen for laks.

5.1 Tradisjonelle vekstkurver

Det er vanlig å anta en eksponensiell vekst over en kortere tidsperiode. Man bruker da en modell der fisken vokser en viss prosent av egen vekt (lengde) per dag. Den eksponensielle vekstkurven er ofte brukt. Dette på tross av at den underestimerer veksten i begynnelsen og overestimerer veksten på slutten av perioden, spesielt hvis perioden er lang.

Andre vekstkurver til å beskrive veksten hos fisk er Bertalaffy, Gompertz, logistisk og powerkurver (Kaufmann 1981).

Disse kurvene baserer seg alle på daglig prosentvis vekst eller spesifikk vekstrate mellom tidspunktene t_0 og t_1 , forkortet til SGR (Specific Growth rate). Imidlertid vil veksten

i prosent dempes med økende størrelse. De nevnte vekstkurvene er modifikasjoner av eksponensiell vekst, og de tar hensyn til at den prosentvise økningen minker med størrelse eller tiden.

Jeg vil ikke gå nærmere inn på alle de nevnte kurvene, men se litt mer på eksponensiell vekst, powerkurven og «termisk vekstkoeffisient».

5.2 Eksponensiell vekst

Det viser seg også at det er vanskeligere å få en god tilpasning når måleserien er «lang». Med «lang» mener jeg både mange måletidspunkt og over en lengre periode. Hver enkelt fisk kan også følge en individuell fleksibel vekstkurve og da kan det være vanskelig å modellere en felles vekstkurve for alle fiskene (Troynikov 1998). Det kan se ut som en individuell laks kan «velge» en vekstkurve eller strategi.

Likevel antar jeg en modell med individeffekter som er tilfeldig fordelt rundt en felles forventning. Jeg diskuterer her kun vektutviklingen, og antar at noenlunde det samme gjelder for lengdeutviklingen.

La oss begynne med en enkel modell med eksponensiell vekst som gradvis gjøres mer kompleks.

For et kortere tidsrom vil en kunne anta at fisken vokser en fast prosent av kroppsvekten per dag (SGR = Specific Growth Rate = fast prosentvis økning av vekten per dag (døgn) i perioden). Det gir samme ligning som eksponensiell vekst og vi får følgende:

$$W(t_1) = (t_0)e^{g(t_1-t_0)} \quad (5.1)$$

$$SGR = (e^g - 1)100, \quad \text{der } g = \ln W(t_1) - \ln W(t_0) / (t_1 - t_0) \quad (5.2)$$

Antar vi at hvert enkelt individ (i) vokser en fast prosent av kroppsvekten per dag i perioden $t_1 - t_0$ får vi ligningen:

$$W_i(t_1) = W_i(t_0)e^{g_i(t_1-t_0)} \quad (5.3)$$

Dette gir én parameter per fisk. Hvis vi ikke har målinger mellom t_1 og t_0 kan tilpasning til dataene over tidsperioden isolert sett se perfekt ut. Men modellen er likevel ikke nødvendigvis tilpasset virkeligheten mellom t_0 og t_1 . Modellen underestimerer som sagt

veksten i begynnelsen av perioden og overestimerer i slutten av perioden. Selv om ikke vi har data mellom t_0 og t_1 er modellen likevel dårlig.

For å få parametre knyttet opp til hver gruppe av fisk ($l = \text{lys}$, $k = \text{kjønn}$ og $m = \text{modning}$) eller for generelt å redusere antall parametre, velger vi heller følgende modell:

$$W_i(t_1) = W_i(t_0)e^{(g_{lkm} + \alpha_i)(t_1 - t_0)} \quad (5.4)$$

Her er α en individeffekt som er normalfordelt med forventning 0 og varians σ_α^2 . Dette gir en «dårligere» tilpasning til dataene, men vi har fått ned antall parametre i hver gruppe av fisk, til to, én for forventningen g_{lkm} og én for standardavviket til α . En utvidelse (flere individeffekter) av dette prinsippet er det jeg bruker ved hjelp av lme-rutinen, bare at jeg velger en tredjegradspolynom som vekstkurve (se kap. 7–10.).

5.3 Termisk vekstkoeffisient

Termisk vekstkoeffisient (Bureau 2000) har vært brukt for å sammenligne vekst hos fisk under ulike temperaturforhold. Effekten av temperaturen blir da tatt hensyn til. På engelsk kalles kurven «Thermic Growth Coefficient», forkortet til TGC. I norsk oppdrettsnæring er den populært kalt V3. Gitt lik temperatur i en periode får vi følgende ligninger:

$$TGC = [(W(t_1)^{1/3} - W(t_0)^{1/3}) / (T \cdot (t_1 - t_0))]100$$

$$W(t_1) = [W(t_0)^{1/3} + TGC(T \cdot (t_1 - t_0)) / 100]^3$$

Der T er temperaturen og $t_1 - t_0$ er antall dager mellom målingene. Mellomregningen går som følger:

$$\begin{aligned} TGC &= [(W(t_1)^{1/3} - W(t_0)^{1/3}) / T \cdot (t_1 - t_0)]100 \\ TGC[T \cdot (t_1 - t_0)] / 100 &= (W(t_1)^{1/3} - W(t_0)^{1/3}) \\ W(t_1)^{1/3} &= W(t_0)^{1/3} + TGC[T \cdot (t_1 - t_0)] / 100 \\ W(t_1) &= [W(t_0)^{1/3} + TGC[T \cdot (t_1 - t_0)] / 100]^3 \end{aligned} \quad (5.5)$$

Har vi ulike temperaturer i ulike perioder bli TGC lik:

$$TGC = 100[(W(t_1)^{1/3} - W(t_0)^{1/3}) / \sum_{i=1}^n (T_i D_i)]$$

Her er T_i temperaturen på dag D_i , og t_0 og t_1 er som tidligere henholdsvis starttidspunkt og sluttidspunkt. I følge Bureau (2000) gir modellen god tilpasning til laksefisk (*Salmonidae*), og har vært mye brukt innen norsk oppdrettsnæring de siste årene, f.eks av Skretting (Hans Petter Endal pers.med.)

5.4 «Powermodellen»

For å ta hensyn til at SGR avtar med størrelsen eller tiden må vi legge til en faktor som reduserer veksten med økende størrelse. En måte å løse dette på er å bruke «powermodellen»:

$$W_i(t_1) = W_i(t_0)e^{(G_i(t)(t_1-t_0))}$$

der

$$G_i(t) = (g + \alpha_i)W_i(t)^{-b}$$

Tar vi hensyn til at veksten er avhengig av temperaturen kan vi tilnærme dette med:

$$G_i(t) = (g + \alpha_i + f_1(cT^2 + dT + e))W_i(t)^{-b}$$

der T er temperatur, f_1 en konstant eller en funksjon og $f_1(cT^2 + dT + e)$ blir ett ledd som gir maksimal vekst ved optimal temperatur. Se (Imsland 2001, s. 24) og referanser i denne for bekreftelse på at SGR lar seg forklare ved regresjonen $SGR = cT^2 + dT + e$.

I tillegg må vi ha en faktor som reduserer veksten på grunn av vintersesongen. Dette kan vi gjøre ved å la faktoren $G_i(t)$ avhenge av daglengden d . Dette gir:

$$G_i(t) = (g + \alpha_i + f_1(cT^2 + dT + e) + f_2(d_t - \bar{d}))W_i(t)^{-b}$$

der \bar{d} er gjennomsnittlig daglengde, f_2 en konstant eller en funksjon og $f_2(d_t - \bar{d})$ blir et ledd som gir vekst avhengig av daglengden, enten lineært eller gjennom en funksjon.

Vi antar en enkel lineær sammenheng mellom $G_i(t)$ og daglengden. Veksten er som nevnt i innledningen også avhengig av om laksen skal gyte inneværende år. Laks som skal gyte kommende år vokser bedre om våren, men veksten stanser utpå ettersommeren/høsten. Veksten er altså avhengig av hvilken fase av kjønnsmodningen fisken er i. Vi må da ha et ledd som trer inn når modningen «starter»:

$$G_i(t) = (g + \alpha_i + f_1(cT^2 + dT + e) + f_2(d_t - \bar{d}) + f_3(M(t)))W_i(t)^{-b}$$

der $f_3(M(t))$ er en funksjon av modning som slår inn på rett tidspunkt.

Hensikten med dette delkapittelet var å peke på mulighetene å til moderere vekstkurver slik at de er bedre tilpasset den biologiske virkeligheten. Powerkurven er ikke nødvendigvis den rette, men er grei til å vise prinsippet. Jeg har kun lagt inn én individeffekt; flere er mulig.

I og med at fiskene i «mitt» forsøk har vokst opp under like temperatur- og lysforhold (sjøvannsfasen) har det ikke så stor hensikt å tilpasse en slik kompleks modell. Men jeg håper at det har kommet klarere fram at laksen ikke har en «fast» veksthastighet, uavhengig av årstid, og at dette har konsekvenser for modellering av blant annet vekstkurver.

De to største fôrfirmaene i Norge (Skretting og Ewos) har utviklet hver sine vekstmodeller (Iversen og Kosmo 2004). Disse formelene er ikke offentlig tilgjengelig (f.eks EGI=Ewos Growth Indeks).

5.5 Forsbergs arbeid

Jeg hadde fortalt Odd Inge Forsberg om at jeg hadde redusert vekst i vintermånende i sjøfasen, men at jeg ikke kunne legge for mye vekt på dette da jeg kun hadde ett kar (en replikasjon). Han spurte da om han kunne bruke dette resultatet sammen med andre vekstforsøk. Sammen med lignende resultater fra andre landbaserte oppdrettsanlegg ble dette da en publikasjon (Forsberg 1995, s. 55–72).

Forsberg modellerer en kurve basert på de samme prinsipper jeg som angir i delkapittelet «Powermodellen». Han lar vekstkurven være avhengig av temperatur, daglengde og størrelse på fisken, basert på «powerkurven». Han definerer vekstrate pr. dag til å være:

$$G_o = 100(\ln W(t_1) - \ln W(t_0)) / (t_1 - t_0)$$

Han bruker videre følgende modell:

$$\ln G = a + b \ln T + d \ln W$$

der G vekstrate per dag, T er temperatur, W fiskens vekt i gram og a , b og c er koeffisienter. Med vekstrate mener han den samme $G_i(t)$ som jeg bruker i «Powermodellen» (men

ganget med 100). Ved hjelp av en multippel regresjon på data fra tabell 3. hos Austreng *et al.* (1987) kommer Forsberg fram til følgende formel:

$$G_E = 0,9T^{0,97}W^{-0,34}$$

der G_E er forventet vekstrate pr. dag, og $a = 0,9$, $b = 0,97$ og $c = -0,34$ innsatt i formelen ovenfor. Modellen angis å være gyldig for temperaturer mellom 4 og 14 grader og fisk mellom 50 og 3000 gram. Basert på $G_E(i)$ og $G_o(i)$ definert ovenfor lages en vekstindeks g_i som et mål på oppnådd vekst i forhold til forventet utfra tabell:

$$g_i = G_o(i)/G_E(i)$$

der i henspeiler på perioden som beregningene gjøres over. Videre introduserer også han daglengde som en ytre forklaringsvariabel for vekstraten. Dette gjør han ved å innføre modellen:

$$\ln G_i = a + b \ln T_i + c \ln \Omega_i + d \ln W_i$$

der G_i vekstrate per dag i periode i , T er midlere temperatur, Ω er gjennomsnittlig daglengde, W fiskens gjennomsnittsvekt i gram og a , b , c og d er koeffisienter. Basert på ulike deler av grunnlagsmaterialet estimeres to vekst rater; G_{NDL} og G_{CL} . Den første tilpasset er naturlig lysperiode (NDL) og den andre er tilpasset kontinuerlig lys:

$$G_{NDL} = 1,4T^{0,68}\Omega^{0,40}W^{-0,38}$$

og

$$G_{CL} = TW^{-0,29}$$

der symbolene har samme betydning som angitt tidligere. Slik jeg har forstått representerte denne artikkelen en steg videre fra tabeller som kun baserte seg på temperatur og størrelse (som tidligere da hadde en vid bruk innen norsk lakseoppdrett).

6 Grafisk framstilling

Det er vanlig å starte en analyse av et datamateriale med grafiske undersøkelser. Derfor gir jeg en kort grafisk presentasjon av deler av materialet, med hensyn på de hypoteser som skal undersøkes. Grafikken gir en idé om modell-tilpasninger som må gjøres. Statistisk analyse er en iterativ prosedyre, der man ser på grafikk, kjører en statistisk metode, ser på residualene, justerer modell eller metode, og ser på residualene igjen, o.s.v.. Man gjentar prosedyren til residualene forhåpentligvis blir som de skal, eller modellen er god nok.

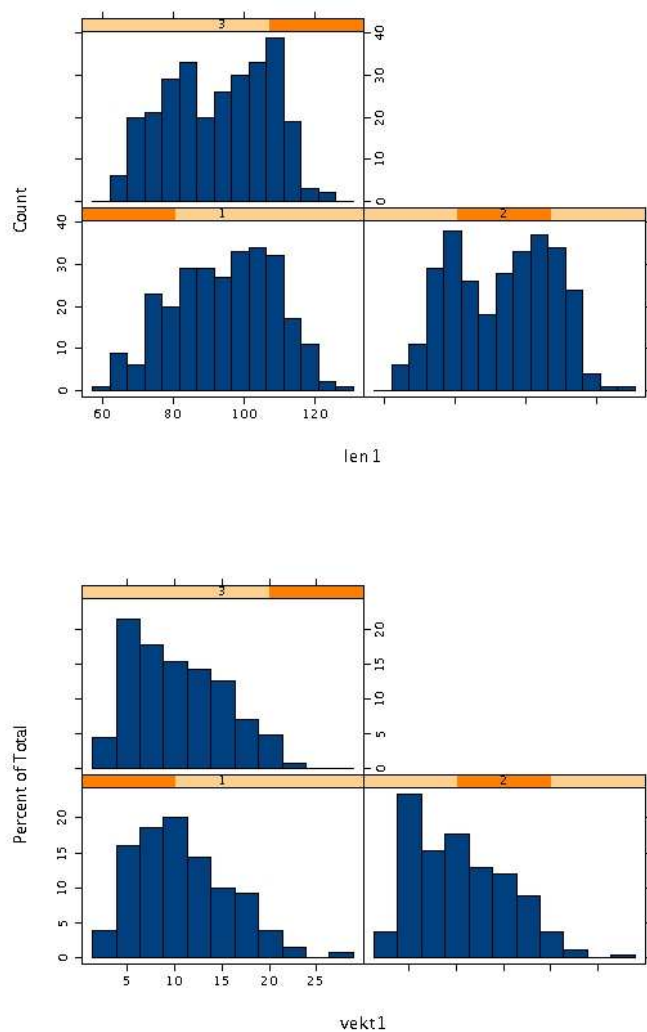
6.1 Fisken ved forsøksstart

Hensikten med dette delkapittelet er å gi en beskrivelse av forsøksfisken ved forsøksstart. For hver lysbehandling er det et histogram av lengde- og vektfordeling, samt spredningsplott av lengde mot vekt. Noenlunde det samme er gjort med hensyn på kjønn og modning. Oppsplittingen er gjort for å vise at det er små eller ingen forskjell mellom forsøksgruppene ved forsøksstart.

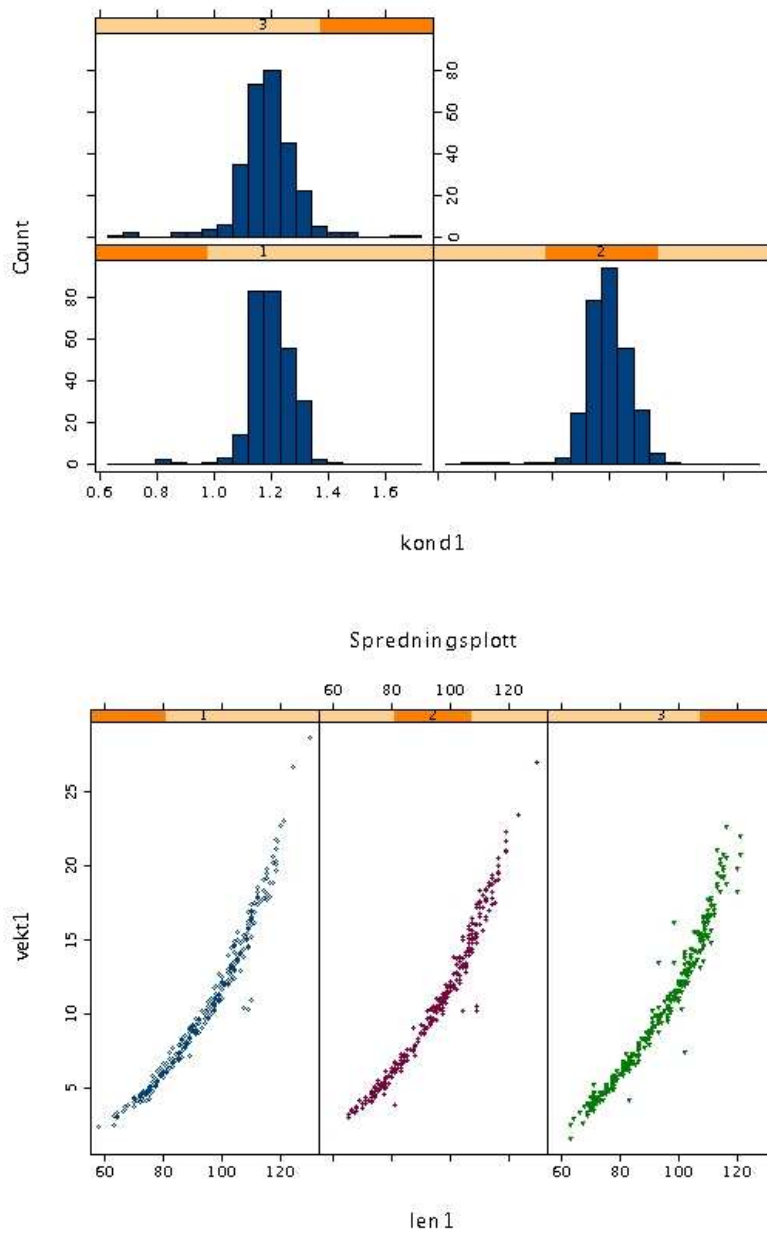
6.1.1 Lys

Visuelt ser det ut til å være liten forskjell mellom fisken som er valgt ut til de ulike lysregimene. For å dokumentere dette ytterligere angir jeg gjennomsnitt og standardavvik for gruppene i tabellen nedenfor.

Vi ser fra tabellen at det er ubetydelige forskjeller mellom fisken som er valgt ut til de ulike lysregimene. Det eneste som er å bemerke, er det er en liten tendens til høyere varians i kondisjonsfaktor, hos fisken som skal ha LD24:0 lysregimet. Jeg antar at det er en tilfeldighet.



Figur 6.1: Histogrammer av lengde i mm og vekt i gram ved forsøksstart. Øverst lengde og nederst vekt. Ett histogram pr. lysbehandling (Gruppe 1 er SNL , gruppe 2 er SNL+B og gruppe 3 er LD24:0).



Figur 6.2: Histogrammer av kondisjonsfaktor ved forsøksstart(Øverst)og spredningsplott av lengde mot vekt (nederst). Delt i lysbehandling (Gruppe 1 er SNL , gruppe 2 er SNL+B og gruppe 3 er LD24:0).

Tabell 6.1: Gjennomsnittlig startvekt, – lengde og – kondisjonsfaktor. Lys1 er SNL, lys2 er SNL+B og lys 3 er LD24:0.

intervall1L	vekt	lengde	K-faktor
Lys1, N=280			
Mean:	10.8	94.3	1.2
Std Dev.:	4.92	14.5	0.074
Lys2, N=294			
Mean:	10.4	93.1	1.19
Std Dev.:	4.9	14.4	0.08
Lys3, N=289			
Mean:	10.3	92.7	1.19
Std Dev.:	4.85	14.5	0.107

6.1.2 Kjønn

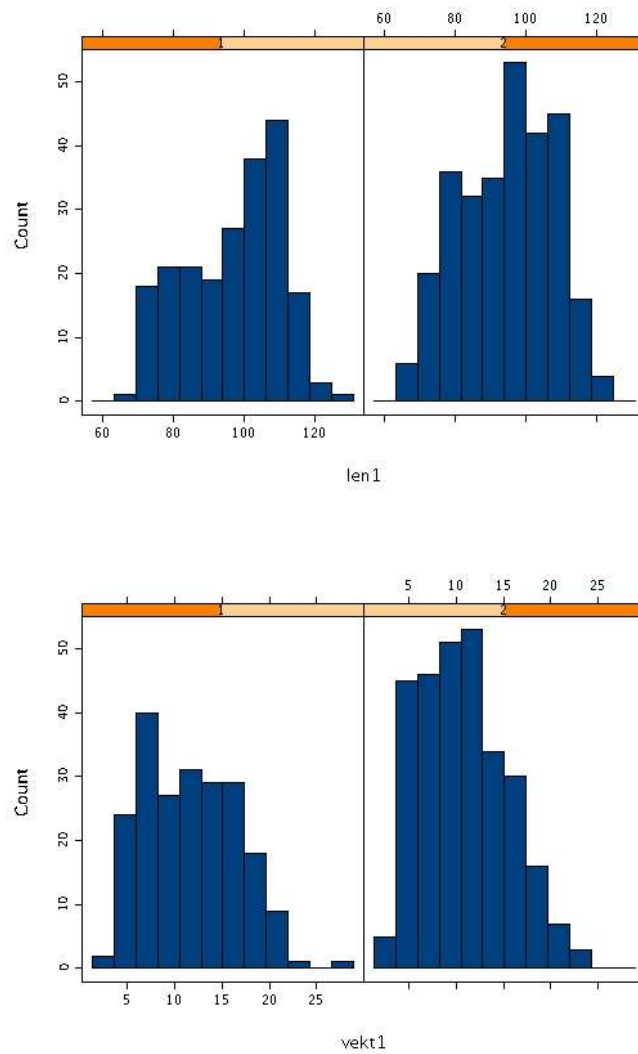
Ufra figur 6.3 på neste side ser det ut at hanner og hunner ikke er vesentlig forskjellige ved start.

6.1.3 Modning

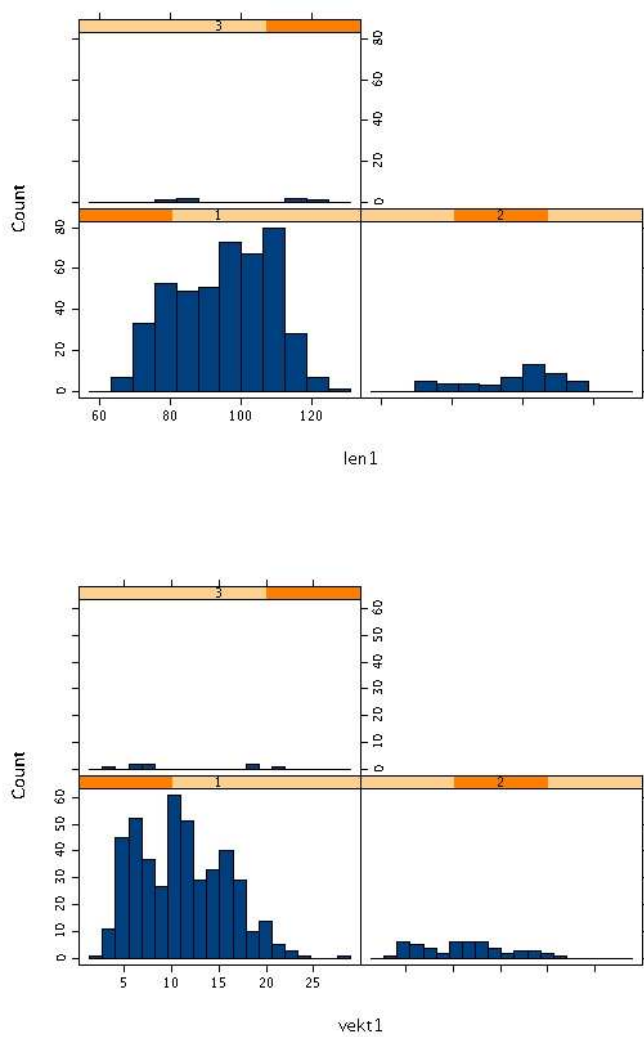
Ufra figur 6.4 på side 62 ser vi at tertene er rekruttert fra et bredt spekter av størrelse. Alternativt formulert: at vi har et tilfeldig utvalg med et bredt spekter av størrelse på terten. De postsmoltmoden kan se ut til å komme fra to størrelseskategorier.

6.2 Ferskvannsfase

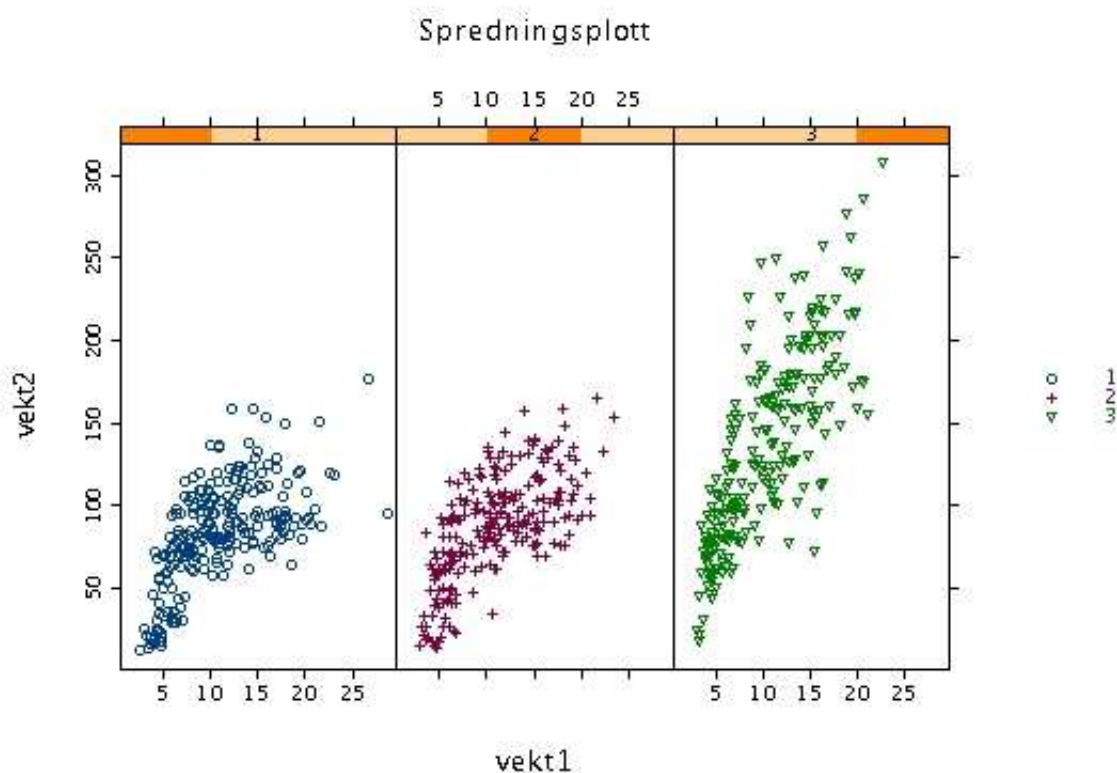
Vi ser av figur 6.5 på side 63 og figur 6.6 på side 64 at det ikke er linær sammenheng og variansen er forskjellig mellom lysgruppene og kanskje økende med startlengde og -vekt. Dette får konsekvenser ved tilpasning av modell. Det er også en antydning til at en gruppe mindre fisk skiller seg ut.



Figur 6.3: Histogrammer av lengde i mm og vekt i gram ved forsøksstart. Øverst lengde og nederst vekt. Ett histogram pr. kjønn. Hanner (1) til venstre og hunner (2) til høyre.



Figur 6.4: Histogrammer av lengde i mm og vekt i gram ved forsøksstart. Øverst lengde og nederst vekt. Ett histogram pr. modningskategori (1=umoden, 2=tert, 3=postsmoltmodne)

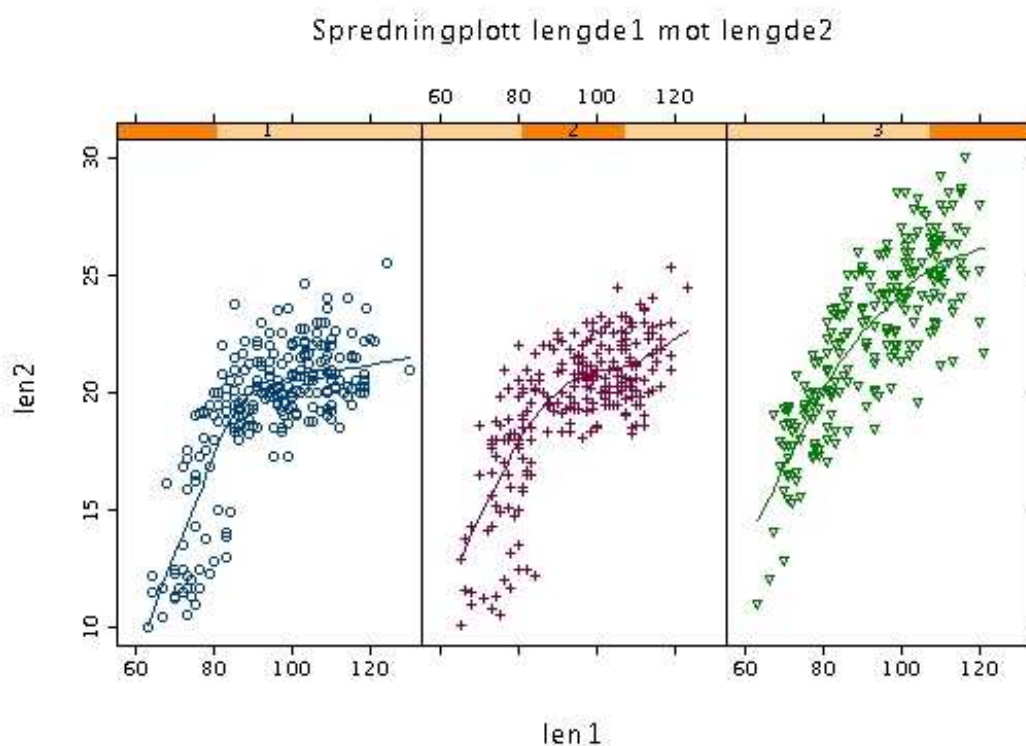


Ett spredningsplott pr lysbehandling.

Figur 6.5: Spredningsplott av vekten 26. mai. 1988 (vekt2) mot vekten 15. september 1987 (vekt1). Plottet er delt opp i tre etter lysbehandlingene. Symboler; 1=SNL : åpen sirkel, 2=SNL+B : kors og 3= LD24:0 : trekant.

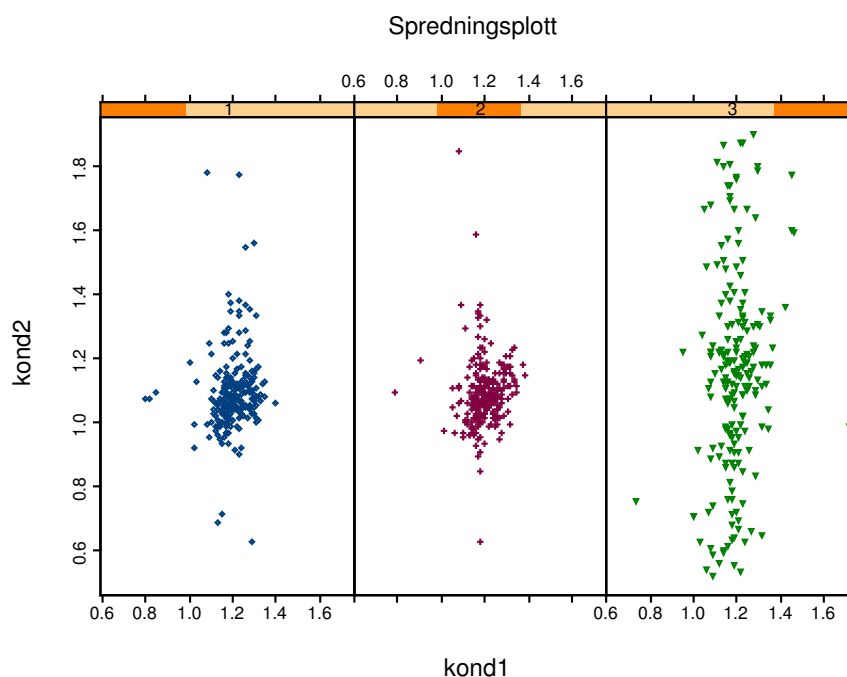
I figur 6.5 ser vi vekten 26. mai 1988 er positivt korrelert med startvekten 15. september 1987. Lysbehandlingen LD24:0 skiller fra de andre gruppene seg ut med hensyn på både «stigningstall», varians og snittvekt 26. mai 1988.

Ut fra figur 6.7 på side 65 ser vi det er vanskelig å tilpasse noe lineær modell for utvikling i k-faktor. Jeg velger derfor en annen tilnærming når det gjelder k-faktor. Videre ser vi av figur 6.7 på side 65 at det ikke er en «pen» lineær sammenheng mellom kondisjonsfaktorene på de to aktuelle tidspunktene. Dette gjør en ANCOVA-modell lite konstruktiv.



Linjer fra ikke parametrisk glatting.

Figur 6.6: Spredningsplott av startlengde mot lengde2 NB! X-akse i mm Y-akse i cm. Symboler som i forrige figur. Linjene representerer glatting med loess-rutinen (opsjoner: span=0.5, førstegrads og symmetrisk).



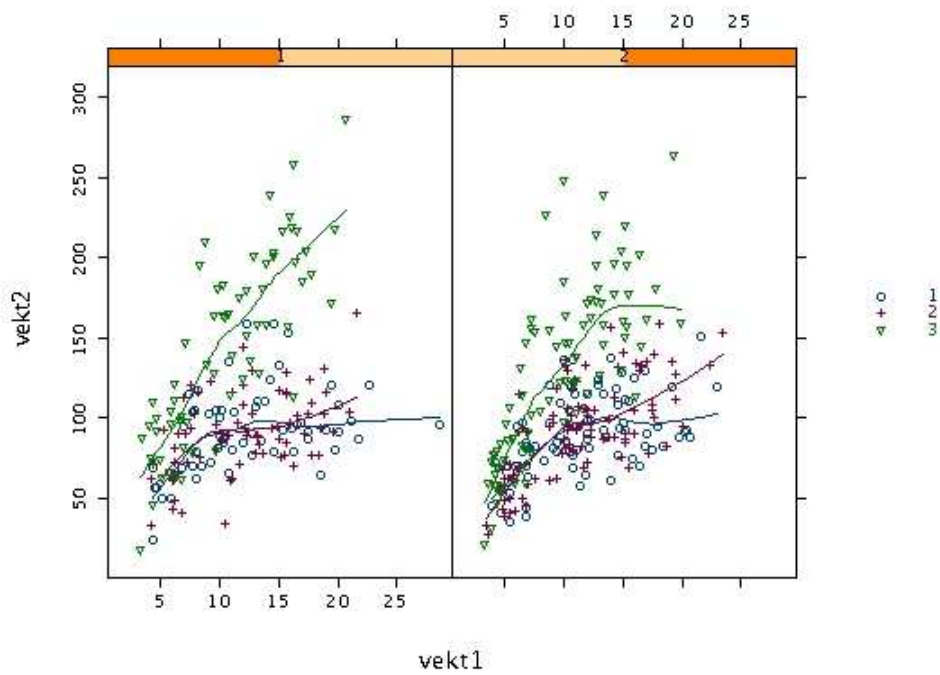
Figur 6.7: Spredningsplott av kondisjonsfaktor1 mot kondisjonsfaktor2. Et plott lysbehandling (1=SNL, 2=SNL+B og 3= LD24:0).

6.3 Smolt

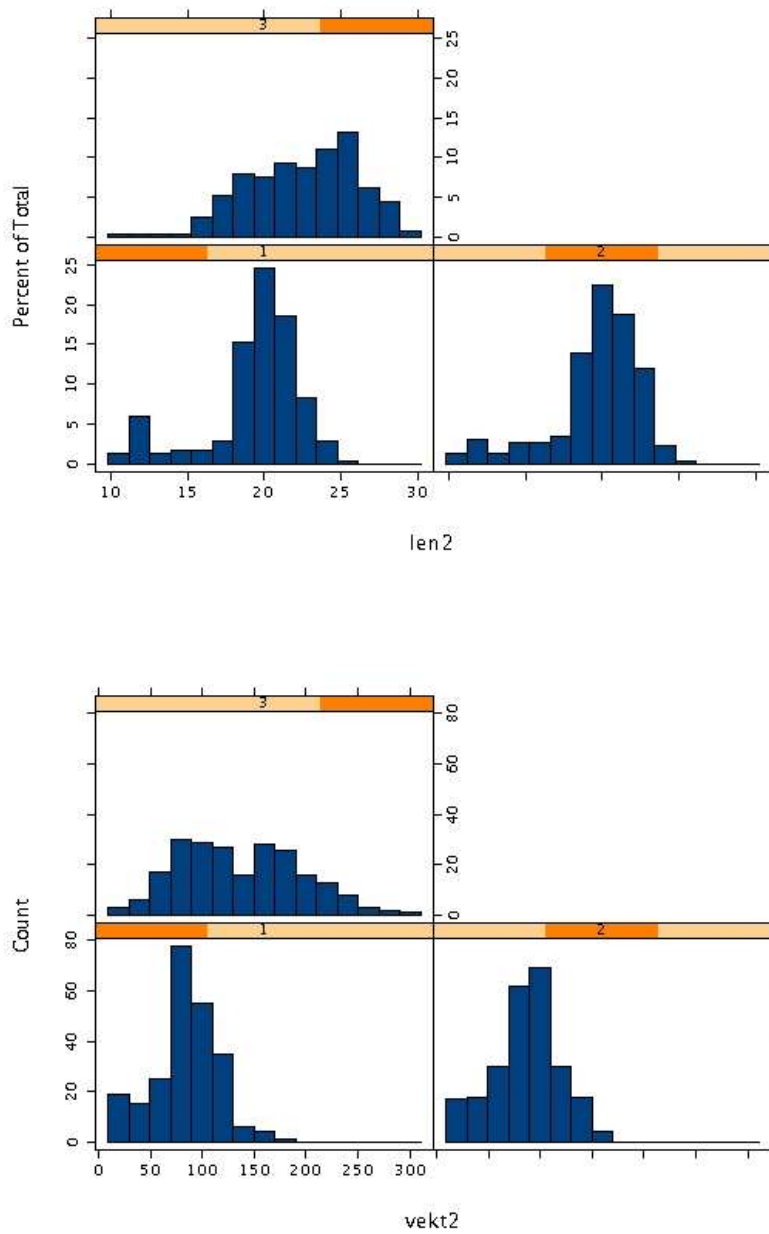
Histogrammene i figur 6.9 på side 67 viser også utvikling av bimodalitet i de to gruppene med SNL.

Vi ser av figur 6.10 på side 68 at i LD24:0 gruppen ikke har samme varians i sammenhengen mellom lengde og vekt, som de to andre gruppene. Dette skal vi se nærmere på når vi trekker kjønn inn i bildet, fordi variansen ser ut til i hovedsak å være knyttet til hunner i LD24:0 gruppen.

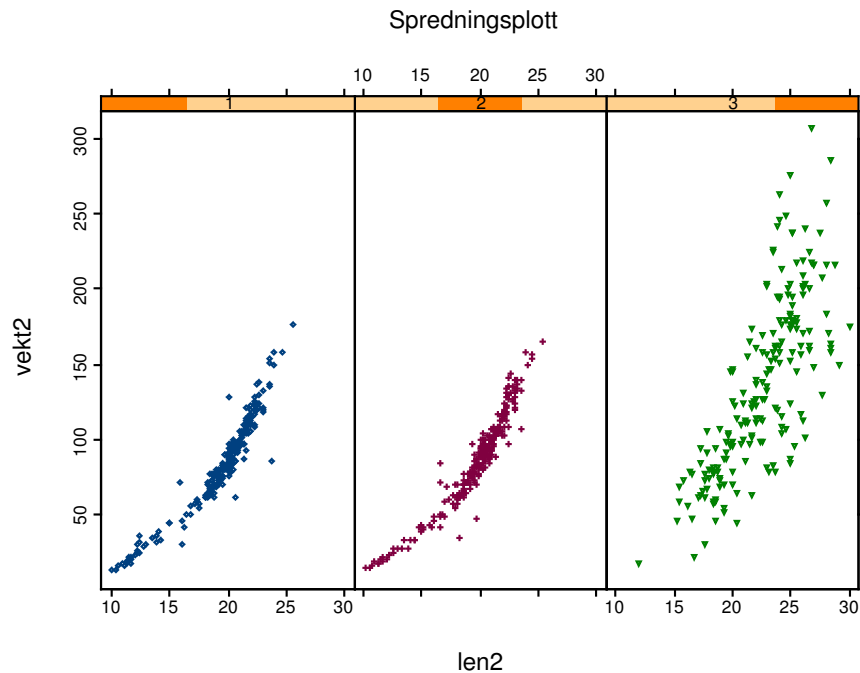
Vi ser i figur 6.11 på side 69 at de korteste fiskene i SNL-gruppene har høyere kondisjonsfaktor. Dette tyder på at disse fiskene ikke er fullstendig smoltifisert.



Figur 6.8: Spredningplott av vekt1 mot vekt2. Plottet er delt i hanner=1 og hunner=2. Ulike symboler etter lysbehandling (åpen sirkel=SNL, kors=SNL+B og trekant= LD24:0). Linjene representerer ikke-parametrisk regresjon ved loess.



Figur 6.9: Histogrammer av smolten (lengde i mm og vekt i gram). Øverst lengde og nederst vekt. Ett histogram pr. lysregime (1=SNL, 2=SNL+B og 3= LD24:0).



Figur 6.10: Spredningsplott av «lengde2» mot «vekt2». Et plott pr. lysbehandling (1=SNL, 2=SNL+B og 3= LD24:0).

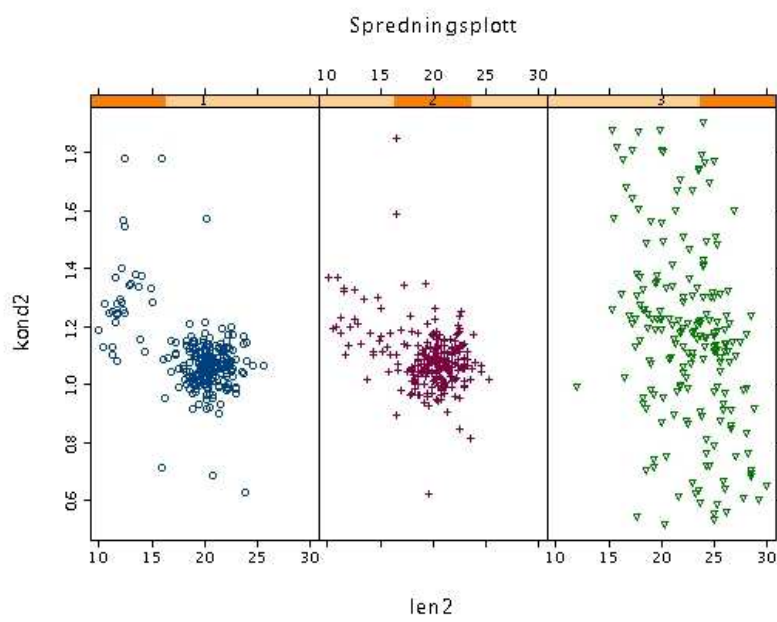
6.4 Samspill: lys og kjønn?

Figur 6.12 på side 70 (øverst) viser at forholdet mellom lengde og vekt ser ganske likt ut for begge kjønn og alle lysregimer. Figur 6.12 på side 70 (nederst) viser at spesielt hunnene innen lysregime LD24:0 har fått en stor spredning.

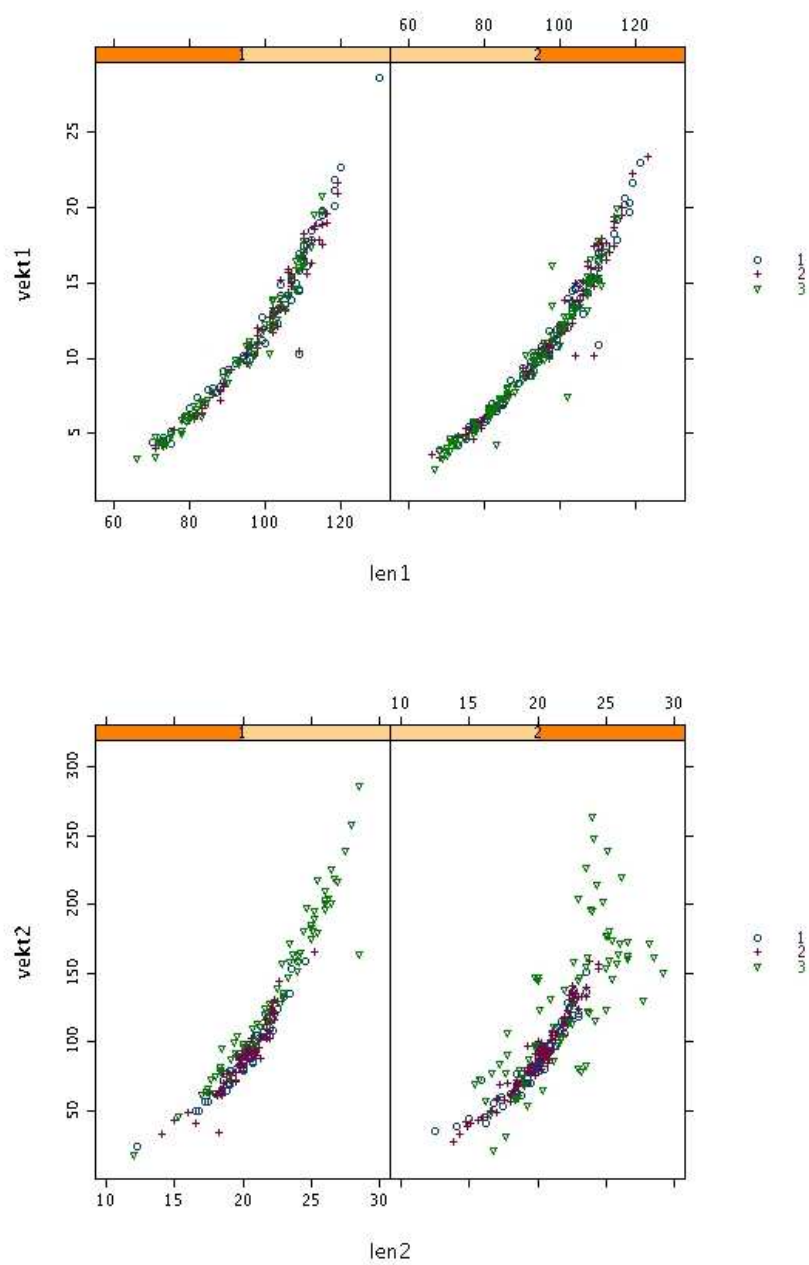
I figur 6.13 på side 71 ser vi at hunnene hos gruppen som fikk lysregime LD24:0 har fortsatt en mye større spredning i forholdet mellom lengde og vekt. Likeså ser vi av figur 6.14 på side 72 (måling 5. og 6.) og figur 6.14 på side 72 (måling 7. og 8.) at denne spredningen varer hele forsøksperioden ut. Vi kan si at dette er en samspillseffekt mellom kjønn og lysregime. Dette blir behandlet nærmere i kap. 9., men det er ikke så lett å trekke noen konklusjon, når det ikke er gjort observasjoner av fiskens atferd.

6.5 Postsmolt

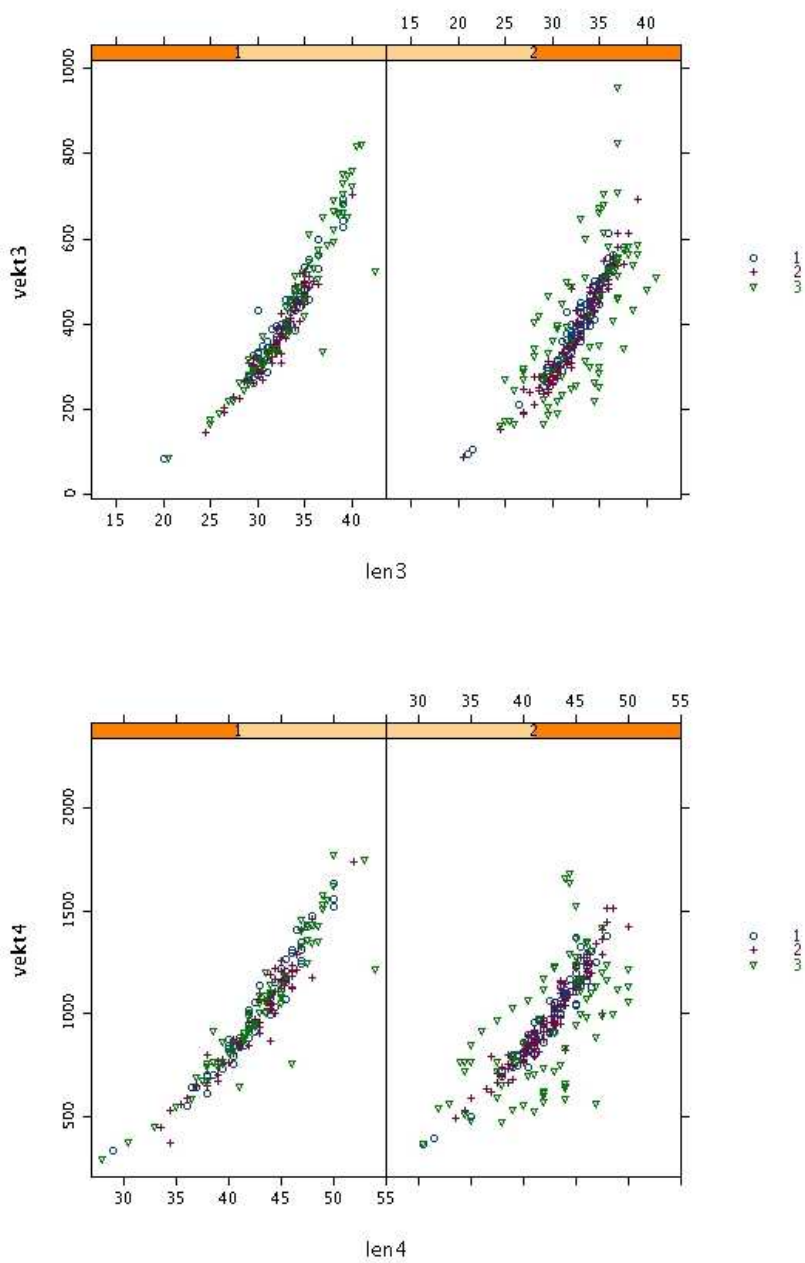
Figur 6.16 på side 74 viser at fisken i de ulike lysregimene har ulik vekst i første del av sjøfasen, når man kompenserer for smoltvekten. Figur 6.17 viser det samme, men bare



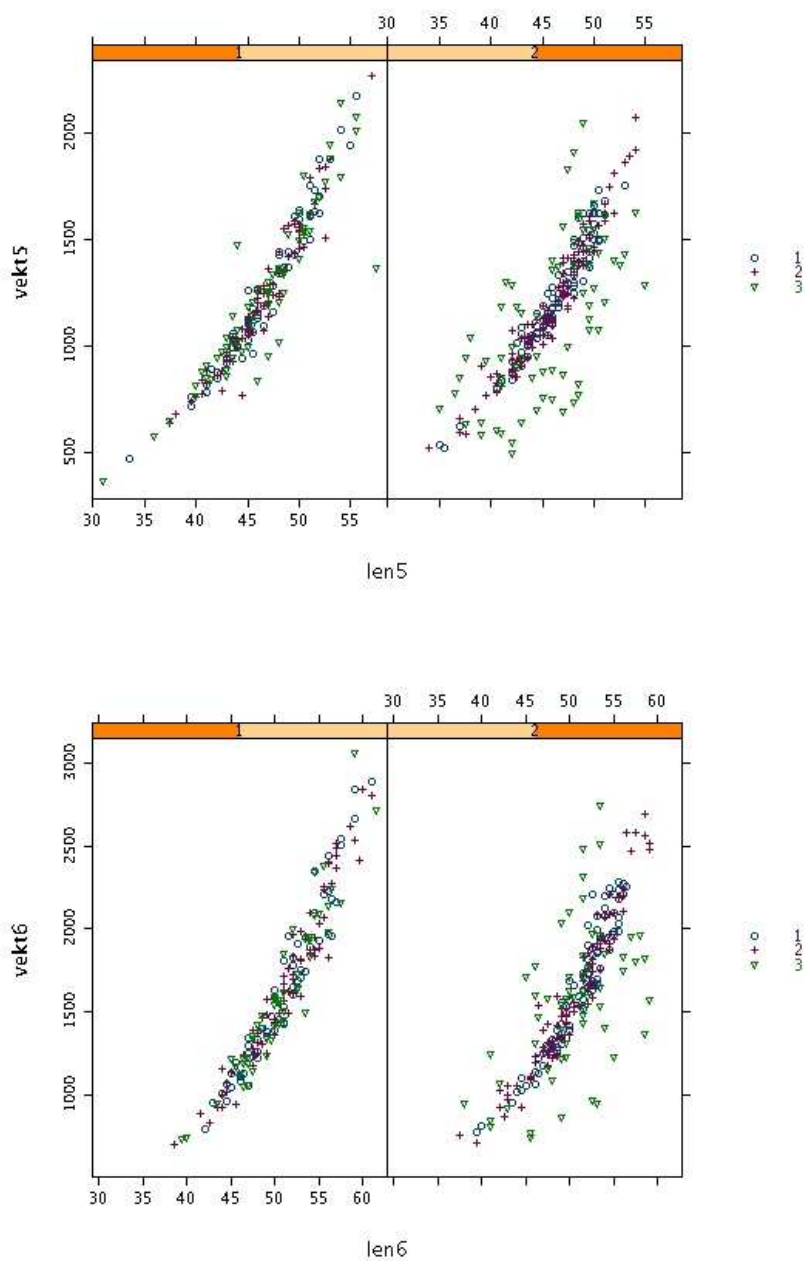
Figur 6.11: Spredningsplott av «lengde2» mot «kondisjonsfaktor2». Et plottet pr. lysbehandling (1=SNL, 2=SNL+B og 3= LD24:0).



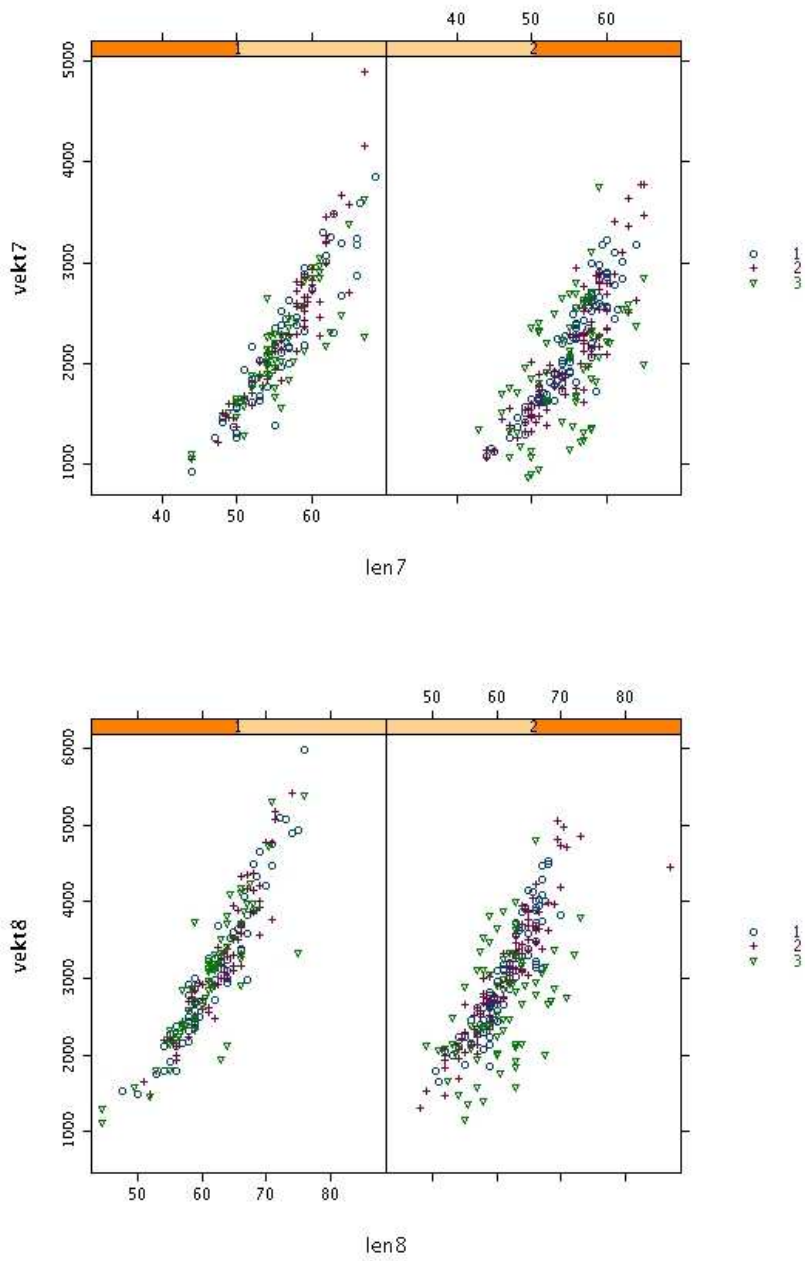
Figur 6.12: Spredningplott av vekt mot lengde. Ulike symboler etter lysbehandling. 1=SNL, 2=SNL+B og 3= LD24:0 (Øverst måling 1. og nederst måling 2.)



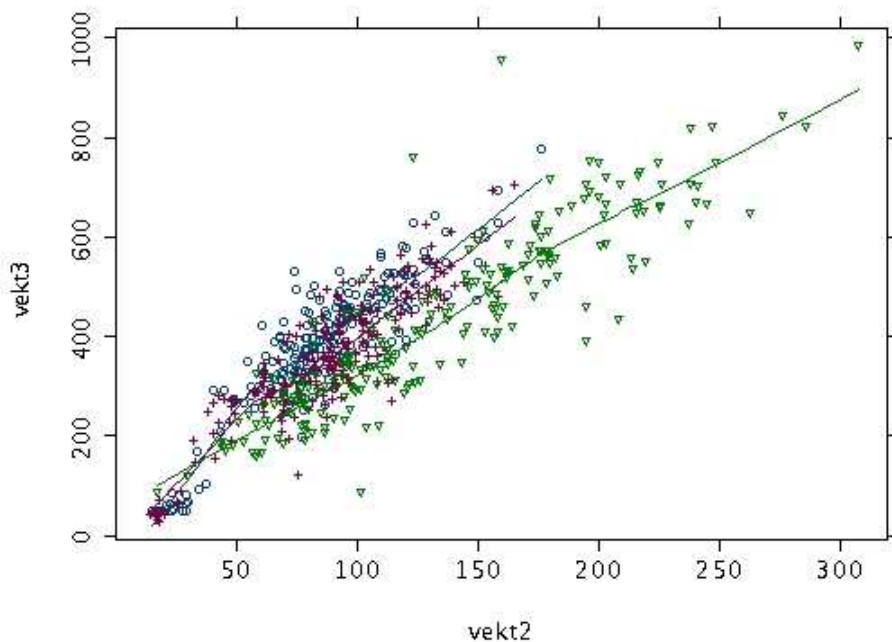
Figur 6.13: Spredningplott av vekt mot lengde. Ulike symboler etter lysbehandling. 1=SNL, 2=SNL+B og 3= LD24:0 (Øverst måling 3. og nederst måling 4.)



Figur 6.14: Spredningplott av vekt mot lengde. Ulike symboler etter lysbehandling. 1=SNL, 2=SNL+B og 3= LD24:0 (Øverst måling 5. og nederst måling 6.)



Figur 6.15: Spredningplott av vekt mot lengde. Ulike symboler etter lysbehandling. 1=SNL, 2=SNL+B og 3= LD24:0 (Øverst måling 7. og nederst måling 8.)

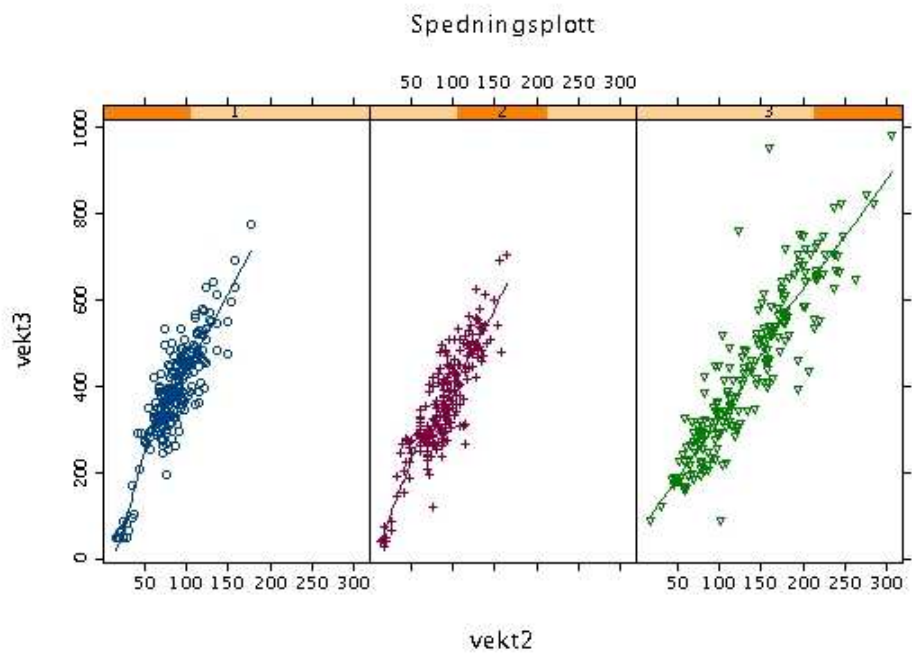


Figur 6.16: Spredningplott av «vekt2» mot «vekt3». Ulike symboler etter lysbehandling (1=SNL, 2=SNL+B og 3=LD24:0).

at her har jeg delt plottet etter lysregime. En lineær modell ser ut til å kunne brukes. Figur 6.18 viser ulik fordeling hos lysregimene m.h.p vekt og utvikling av bimodalitet.

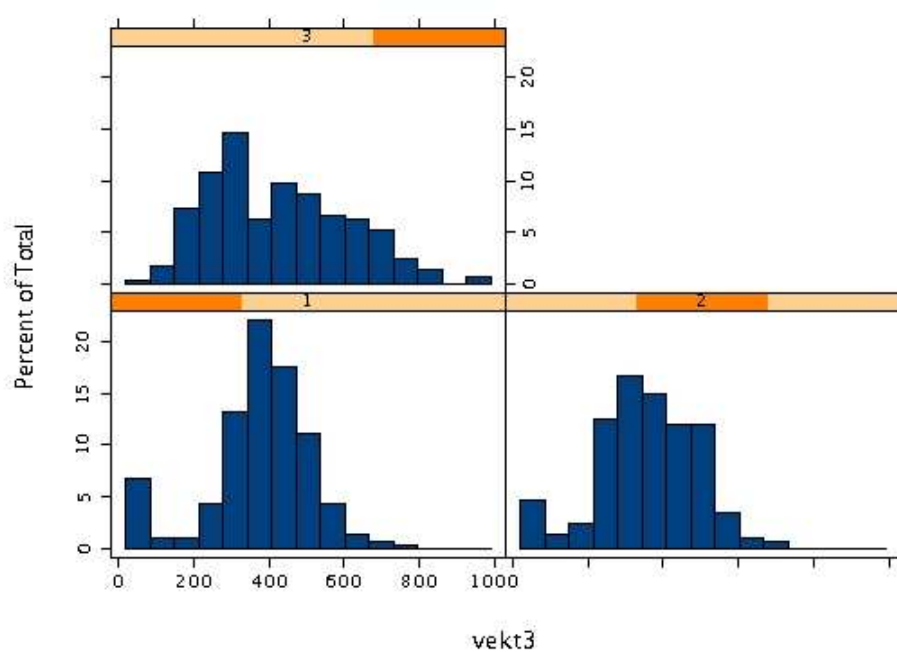
6.6 Vekstforløp

Ufra figur 6.19 på side 77 ser vi at variansen er økende med tiden. Til en viss grad er det også mulig å se at store fisker forblir store og likeså for små. Dette fenomenet kalles

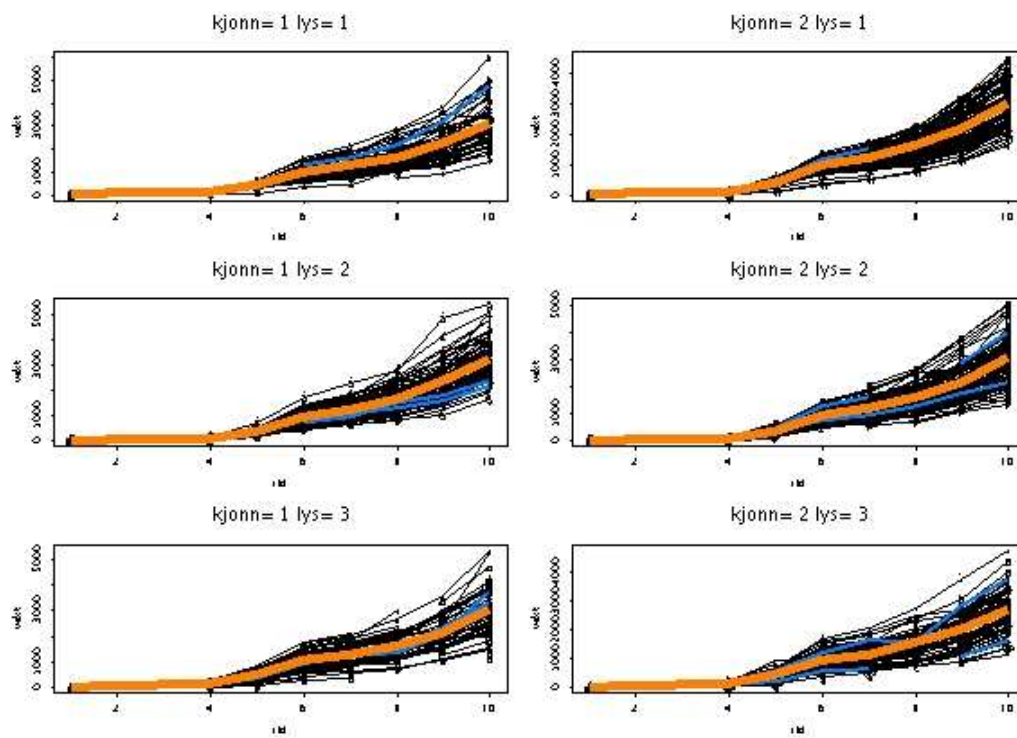


Ikke parametriske regresjon ved loess.

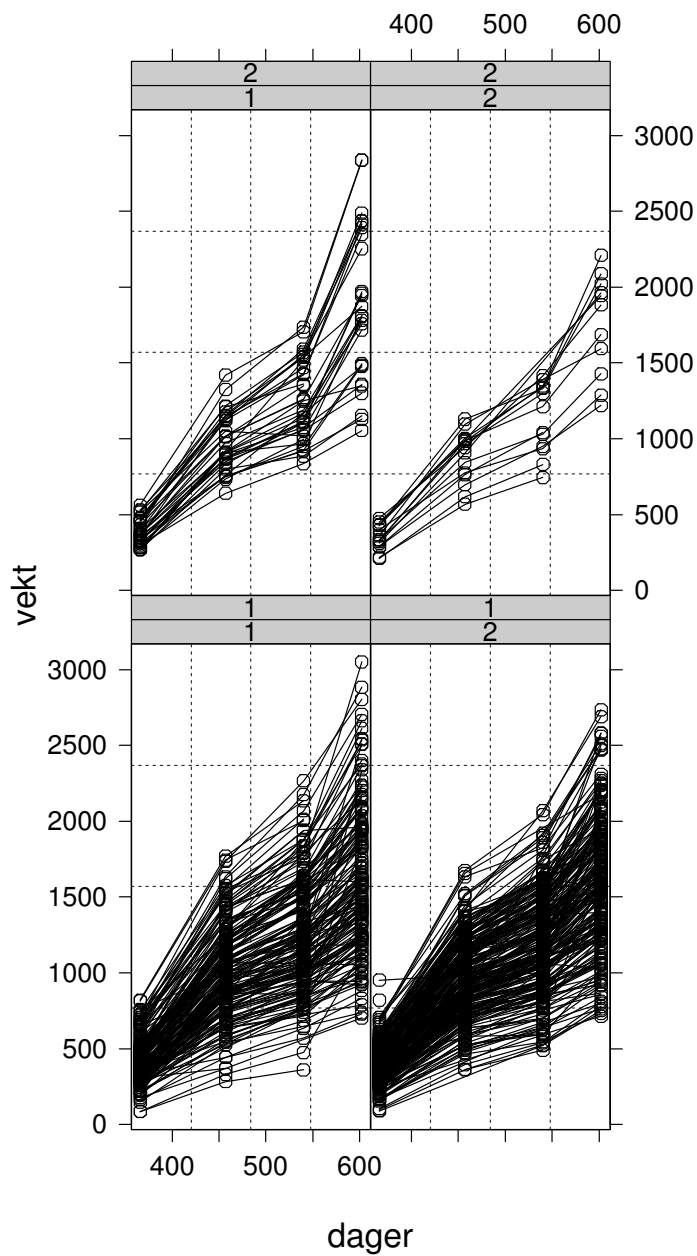
Figur 6.17: Spredningsplott av vekt3 mot vekt2 . Linjene representerer loess (). Plottet er delt opp i tre, etter lysbehandling. 1=SNL, 2=SNL+B og 3= LD24:0.



Figur 6.18: Histogram av vekt3. Histogrammet er delt opp i tre, etter lysbehandling. 1=SNL, 2=SNL+B og 3= LD24:0.



Figur 6.19: Plott av vekt-utviklingen pr. individ. Plottet er delt opp i seks, etter kjønn (1= hanner og 2=hunner) og lysbehandling (1=SNL, 2=SNL+B og 3= LD24:0).



Figur 6.20: Plott av en kurve for hver enkelt fisk (umodne hanner; nederst til venstre, umodne hunner; nederst til høyre, modne hanner; øverst til venstre og modne hunner; øverst til høyre).

7 Statistiske metoder

7.1 Valg av statistiske metoder

7.1.1 Type studier

Det er viktig å skille mellom «data-mining», observasjonsstudier og kontrollerte forsøk med klare hypoteser.

- Data-mining
 - leter generelt etter sammenhenger
 - ofte ingen klare hypoteser
- Observasjonsstudier
 - ikke kontrollerte forsøk
 - delvis klare hypoteser
- Kontrollerte forsøk
 - mest mulig kontrollerte forsøk
 - klare hypoteser

Med delvis klare hypoteser mener jeg vidt formulerte hypoteser eller vage formuleringer, altså dermed muligens mange hypoteser.

Man skiller mellom disse situasjonene når man rapporterer resultatene. Ellers risikerer man for eksempel å rapportere artefakter som ordinære resultater. Type situasjonen vil også kunne påvirke valg av signifikansnivå.

I vårt «forsøk» er vi i et grenseland. Dette fordi vi i utgangspunktet startet med et kontrollert «forsøk» med tre lysbehandlinger i ferskvannsfasen, og relativt klare hypoteser.

Deretter ble det bestemt at fisken kunne settes på sjøvann, for å se om man fant spennende resultater ved å følge opp forsøket videre. Hypotesene blir da flere og mer uklare. Samtidig blir alle fiskene plassert i ett kar, og det er dårligere oppfølging av forsøksbetingelsene. Dette gir mer preg av en observasjonstudie.

Det ideelle ville vært to eller flere paralleller også i sjøvannfasen. Man ville da være mer sikker på at observasjoner ikke skyldtes kareffekter som kan være et problem når vi ser på det generelle vekstmønsteret til alle grupper av fisk samlet. På den annen side vil det være en avveining mellom mest mulig realistiske driftsforhold (store enheter) og antall paralleller.

Noen av problemstillingene i sjøvannfasen var bestemt på forhånd. Man ønsket å undersøke forskjeller i vekstmønster mellom lysbehandlingene, mellom hanner og hunner, samt mellom laks som modnes etter ett år i sjøen og seinere. Men det var også ønske om å få kartlagt sider ved veksten som ikke var klart definert.

Leter man generelt etter sammenhenger i datasettet blir dette «data-mining».

Finner man noe spennende, kan dette eventuelt følges opp med nye forsøk hvor man undersøker om resultatene er repeterbare. De individuelle vekstkurvene kan være interessante, selv om vi nå først og fremst ser etter forskjeller mellom de angitte laksegrupper.

7.1.2 Konstruksjon av modell

Statistisk metode bestemmes ut i fra materialets beskaffenhet og hvilke problemstillinger man vil undersøke. Det er kombinasjonen som er avgjørende. Både forsøksoppsettet, måle metodene, forsøkspopulasjonen og problemstillingene er i prinsippet med på å avgjøre modellen.

- Forsøksoppsettet
 - gir behandlinger, grupper og antall målinger
- Måle metodene
 - gir målefeil og eventuelt feilklassifisering
- Forsøkspopulasjonen
 - gir varians og eventuell individeffekt
- Problemstillinger

- gir hypoteser og styrer hva som er responser og forklaringsvariabler

Vi lager en modell med responser, forklaringsvariabler og ulike feilledd. Videre antar vi gjerne noe om fordeling, uavhengighet, samspill og struktur i eventuelle kovariansmatriser. Uansett må kompleksiteten til modellen stå i et rimelig forhold til datasettet man holder på med.

Statistiske kunnskaper til personen som skal bearbeide datasettet vil i praksis være avgjørende. Ofte vil selve forsøket være utformet slik at vi kan bruke en godt kjent statistisk metode, f.eks. toveis ANOVA.

Det som er spesielt i vårt forsøk er gjentatte målinger på samme individ. Det er én måleserie for lengde og vekt for hver fisk. De laksene som overlevde til forsøkslutt er målt åtte ganger i løpet av to år.

Dette medfører at vi kan velge mellom flere statistiske metoder, noe som diskuteres i neste avsnitt.

7.1.3 Aktuelle metoder

Dataene våre kan betraktes som tidsserie- eller paneldata, repeterte målinger eller longitudinelle data. Data med gjentatte målinger på samme individ kan og høre hjemme under betegnelsen lineære blandede modeller eller «Mixed effects models». Disse begrepene brukes ikke helt entydig, og derfor gir jeg noen få stikkord om bruk av begrepene:

- Tidsserie
 - Gjentatte målinger på samme individ, med en eller få lange tidsserier. Autokorrelasjon vil være helt sentralt.
- Paneldata
 - Brukes om mange korte tidsserier. Ofte autokorrelasjon.
- Repeterte målinger
 - Brukes ofte om vekslende behandlinger over tid (randomisert), for eksempel i medisinske forsøk. Ble før brukt synonymt med tidsrekker og longitudinelle data.
- Longitudinelle data
 - Brukes generelt om gjentatte målinger på samme individ i tid. F.eks. vekstdata.

- Vekstmodeller
 - Longitudinelle modeller ved vekt på kurvasjonen. Kalles ofte «growth models» med vekt på polynomiske modeller/tilpasninger.
- Lineære blandede modeller (LME)
 - Blanding av «faste»- og «variens»-komponenter. Preget av pseudoreplikasjoner på samme gruppe eller individ. Med pseudoreplikasjoner menes for våre data flere målinger spredt i tid på samme individ (se nedenfor).
 - Jeg bruker metoden med faste kurver og en utstrakt bruk av individeffekter som da er varienskomponentene (se avsnitt 7.5 om LME).

Pseudoreplikasjoner kan defineres som replikater som ikke er statistisk uavhengige eller henspeile på misbruk av statistiske metoder p.g.a. nevnte forhold (Hurlbert 1984). I data-settet «vårt» er dette en svært aktuell problemstilling, som jeg forsøker å løse ved hjelp av multivariate metoder.

Andre aktuelle metoder er som nevnt tidligere (i avsnitt 1.6) er henholdsvis ANOVA, ANCOVA, MANOVA/MANCOVA, logistisk regresjon og diskriminantanalyse.

7.1.4 Valgte metoder

I dette avsnittet går jeg inn på hvilke metoder jeg har valgt og hvorfor. Jeg har valgt å bruke følgende tilnæringer til dataene:

- ANCOVA
- MANCOVA/MANOVA
- Lineære blandede modeller

I Splus har jeg løst dette ved bruke følgende rutiner, nevnt i samme rekkefølge:

- gls-rutinen
- manova-rutinen
- lme-rutinen

For å se på respons i et tidspunkt har jeg brukt ANCOVA. I forhold til ANOVA er dette en betydelig forbedring i mange situasjoner, men ikke i alle.

Der jeg ønsker å se på flere tidspunkt sammen har jeg brukt henholdsvis lineære blandede modeller (LME) og MANOVA/MANCOVA. Jeg har imidlertid tonet ned bruken av LME, da metoden ikke tilfører noe rent biologisk utover det jeg finner med MANOVA/MANCOVA. Bruken av LME er vist i appendiks D. Jeg skriver likevel litt om teorien til metoden i dette kapitlet da jeg mener metoden har noe for seg ved analyse av flere måletidspunkt enn det jeg har valgt å bruke.

En hovedproblemstillingen er altså å velge metode for å beskrive dataene, enten med vekstkurver eller med MANOVA. Hvilken metode man bruker vil blant annet avgjøres av antall målinger per individ og eventuelt antall parametre i vekstkurven.

Forskjellen ligger også i hvordan forventningene estimeres. Vi kan forenklet si at MANOVA er punktestimering, mens vekstkurver dreier seg om kurvetilpasning. Ett annet poeng i valget mellom MANOVA og vekstmodeller er at med MANOVA kan vi se på lengde og vekt sammen. Det er ikke så lett når vi bruker vekstkurver. MANCOVA er litt annerledes, men det er også knyttet vansker til denne metoden. I praksis blir det avveininger i forhold til datasett, og hva man er ute etter.

Noen av tilnærmingene kan oppfattes som forsøk på å nærme seg Diggles modell for longitudinelle data (Diggle 1990). Hans modell er fleksibel, og man kan der velge mellom MANOVA eller vekstkurver. Man kan også modellere feilledet på ulike måter, slik at man kan ta hensyn til individeffekter og eventuell autokorrelasjon.

En annen fordel med Diggles modell er at forskjellige fisk kan være målt på ulike tidspunkt, noe som er viktig ved bruk av «ny måleteknologi». Dette ser vi mer på under en utførlig beskrivelse av denne modellen.

Disse tilnærmingene vil jeg utdype nedenfor, men først litt om logistisk regresjon. For å begrense oppgavens omfang utgår bruk av logistisk regresjon fra oppgaven.

Logistisk regresjon

Dette er i prinsippet interessant fordi vi kan ønske å predikere modning eller ikke for hvert individ utfra vekt og lengde, som er kontinuerlige variabler.

Man kan også rangere MANOVA, logistisk regresjon og diskriminantanalyse på følgende måte: Dersom man ikke får utslag på en MANOVA, vil ikke diskriminantanalyse ha noe

for seg. Logistisk regresjon kan oppfattes som ett steg midt mellom MANOVA og diskriminantanalyse. Dette er selvfølgelig en forenkling. Man vil også som sagt få «bedre» resultater dersom ikke nedre modal var fjernet (se forord 1.3).

Jeg vil bare kort nevne at jeg mener å ha oppdaget samme fenomen som beskrevet av Hauck og Donner (1977): ikke samsvar mellom Wald-observatoren og differansene til deviansene.

Wald-observatoren kan ha en del uønskede egenskaper med hensyn på å teste enkeltparametre i logistisk regresjon. Observatoren minker til null når parameterestimatet beveger seg vekk fra null. Styrken går mot α i alternativ som er langt fra null i store utvalg. Dette er i Venables og Ripley (2003, s. 197–198) betegnet som et lite kjent fenomen.

7.1.5 Spesielle problemstillinger

Uavhengighet

Alle metoder jeg bruker forutsetter uavhengighet mellom hver enkelt fisk. Dette er ikke tilfelle da fiskene går i samme kar og konkurrerer om beste plassering i karet og om fôr. Vi kan gjøre lite med det og antar likevel at tilnærmingen er rimelig god.

Antar vi multinormalfordelte data får avhengigheten mellom ulike fisker følgende konsekvens:

La vektoren \mathbf{y} inneholde alle målingene (i våre data: pn). Antall observasjoner på samme individ betegnes p og antall individ er n . Da er vektoren \mathbf{y} multinormalfordelt: $N_{np}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, og får vi følgende sannsynlighetstetthet:

$$f(\mathbf{y}) = \frac{1}{(2\pi)^{np/2} |\boldsymbol{\Sigma}|^{1/2}} e^{-(\mathbf{y}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y}-\boldsymbol{\mu})/2}$$

Alle fiskene går i ett kar og vi har bare én multinormalfordelt observasjon. Ser vi litt nærmere på $\boldsymbol{\Sigma}$ så kan vi si at langs blokkdiagonalen i denne ligger hver fisk sin $\boldsymbol{\Sigma}_i$. Samspillet mellom fiskene ligger utenfor blokkdiagonalen.

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \dots & \boldsymbol{\Sigma}_{1n} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & \dots & \boldsymbol{\Sigma}_{2n} \\ \vdots & \vdots & \ddots & \\ \boldsymbol{\Sigma}_{n1} & \boldsymbol{\Sigma}_{n2} & \dots & \boldsymbol{\Sigma}_{nn} \end{bmatrix}$$

Vi antar altså alle leddene utenfor blokkdiagonalen er $\mathbf{0}$. Rimeligheten av antagelsen er avhengig av fôring, plassforhold, sosiale forhold etc. Se Imsland (2001) for simulering av hierarki på piggvar.

Jeg vil sterkt poengtere at fiskene er tilfeldig valgt ut i starten av forsøket. Vi betrakter derfor målingene videre på hvert individ som uavhengige observasjonsvektorer, der alle fiskene utsettes for noenlunde samme miljø.

Antar vi da n uavhengige y_i er blir simultansannsynlighetstettheten;

$$\begin{aligned} f(\mathbf{y}_1 \cdots \mathbf{y}_n) &= \prod_{i=1}^n \left\{ \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} e^{-(\mathbf{y}_i - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{y}_i - \boldsymbol{\mu}) / 2} \right\} \\ &= \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} e^{-\sum_{i=1}^n (\mathbf{y}_i - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{y}_i - \boldsymbol{\mu}) / 2} \end{aligned} \quad (7.1)$$

der kovariansmatrisen Σ nå er lik for hver fisk.

Ubalansert datasett

Datasettet er ubalansert med hensyn på antall modne og ikke modne. Dette kan gi problemer i flerveis-ANOVA og -MANOVA basert på klassiske metoder. Ved bruk av likelihood-estimering og loglikelihood-forholdstesten vil dette ikke i samme grad være noe problem (Searle *et al.* 1992).

Uteliggere og Robuste metoder

I datasettet var det i utgangpunktet mange uteliggere. Disse har jeg rensket ut for hånd basert på biologisk kunnskap (se avsnitt om kvalitetesikring 4.7). Dette kunne ha vært gjort mer metodisk ved å kombinere biologi og statistikk (Se Pan og Fang (2002) eller Wilcox (2001)). I Pan og Fang (2002) angis to ulike tilnærminger for problemet med uteliggere:

- «Statistisk diagnostikk»
 - Finne observasjoner som avviker fra modellen.
 - Identifisere innflytelsesrike observasjoner.
- «Robust statistikk»

- Bruk av statistiske metoder som minsker effekten av uteliggere.
- Robuste metoder overfor små og moderate forstyrrelser.

I forlengelsen av at datasettet inneholder en del uteliggere ville det vært naturlig å se på ulike mer robuste metoder (Se både Pan og Fang (2002) og Wilcox (2001)).

Dette har jeg ikke gjort fordi det var en del av planen og avtalen for dette hovedfaget at jeg ikke skulle se på disse problemstillingen.

Jeg har heller satset på at dette arbeidet som jeg gjorde under hovedfaget i biologi no-enlunde holder mål. Bruk av robuste metoder har ofte med dårlig kvalitetsikrede data å gjøre. Jeg har tro på at dataene er gode nok til at dette ikke er det mest presserende problemet.

Bootstrapping

En alternativ måte å få konfidensintervaller for estimatene på er å bruke bootstrapping. Det innebærer å gjøre mange utvalg fra dataene med tilbakelegging, for så å reestimere aktuelle parametre fra hvert enkelt utvalg. Ved så å betrakte prosentilene fra denne prosedyren får man ett innblikk i usikkerheten til estimatene (Crawley 2002, s. 195–201). Vi antar her at dataene avspeiler den sanne fordeling i populasjonen.

En alternativ metode er paramerisk bootstrapping, da estimerer man fordelingsparametre fra populasjonen. Etterpå trekker man fra en «populasjon» med de estimerte fordelingsparametrene, og følger videre samme prosedyre som ovenfor (Scheiner og Gurevitch 2001, s. 273–288). Man tilnærmer seg da problemet ved en noe annerledes prosedyre, som i sitt vesen kan ligne noe på bootstrapping av residualer. Begge prosedyre antar riktig spesifisering av enten fordeling eller modell.

Om bootstrapping i linær regresjon skriver (Venables og Ripley 2003, s. 164) i «Modern Applied Statistics with S» følgende: «If the errors are close to normal, the standard teory suffices.» Her myntet på minste kvadraters metode.

Dette har altså først og fremst noe for seg hvis det er grunn til å tvile på at dataene følger den fordeling som forutsatt. Men jeg er inneforstått med at bootstrapping kunne vært et nyttig supplement.

Manglende data

Mange fisker «døde» eller «forsvant» underveis i ekperimentet, og gir dermed manglende data. På de fiskene som døde underveis har man ikke registret kjønn noe som lett kunne vært gjort. Dette gjør at vi mister mange fisker i forhold til å se på kjønnsforskjeller. Vi har også manglende data på grunn av fisker uten PIT-nummer og av utrenskningen jeg har gjort av uteliggere.

Har kunne det altså vært aktuelt å bruke EM-algoritmen for å «rekonstruere» manglende data (Johnson og Wichern 2002, s. 252–256). Algoritmen baserer seg da på at det er tilfeldig hvilke data som mangler.

Å gjøre dette er først og fremst aktuelt dersom dataene er for få til at vi kan estimere det vi er ute etter. Algoritmen kan imidlertid også brukes på sensurerte data.

7.2 Om estimering og testing

Mitt fokus er ikke først og fremst testing, men estimering og konfidensintervaller. Vi kan generelt si at test-resultater bare angitt med signifikans eller ikke, er mindre informativ enn å oppgi p-verdiene. Lager man konfidensintervaller vil det igjen være mer informativt. Jeg foretrekker en kombinasjonen av p-verdier og konfidensintervaller.

Statistikk er en prosess med prøving og feiling for å finne en god modell. Man sammenligner ulike modeller og prøver å finne en modell som beskriver datene godt uten å være for komplisert.

For dette formålet bruker jeg sannsynlighetkvotetesten ¹, p-verdier og diagnostiske plott som veiledning.

Vi ønsker en modell som er enkel, men godt tilpasset dataene. Mange parametre vil generelt gi bedre tilpasning til data. Men ved bare marginalt bedre tilpasning med de ekstra parametrene sier vi at de ekstra parametrene ikke har «forklaringskraft».

¹På norsk brukes også loglikelihood-forholds-testen og på engelsk «loglikelihood-ratio-test». Jeg bruker de norske begrepene om hverandre.

7.2.1 Sannsynlighetskvotetesten

Sannsynlighetskvotetesten foregår slik at man starter med en kompleks modell og så forenkler. Dersom loglikelihood-forholdstesten ikke gir en lav p -verdi forsetter man forenklingen. Dersom de parametrene som fjernes ikke er av noe betydning, vil forholdet mellom loglikelihoodene være χ^2 -kvadratfordelt med antall frihetsgrader lik antall fjernede parametre.

En høy p -verdi betyr at den enklere modellen er omtrent like god som den mer komplekse. Får vi en lav p -verdi så beholder vi den komplekse modellen. Vi antar da at denne faktisk forklarer dataene betydelig bedre.

Dette er ikke ett verktøy man kan bruke helt blindt og ukritisk. Det finnes en del korreksjoner til metoden der man «straffes» for antall parametre p i modellen, f.eks. informasjonskriteriet AIC:

$$AIC = -2\log L(\mathbf{b}) + 2(p + 1) \quad (7.2)$$

Der \mathbf{b} er maksimumlikelihoodestimatet med p parametre.

Differanser mellom devianser brukes synonymt med loglikelihood-forholds-testen. I prinsippet går testen ut på å vurdere forholdet mellom likelihoodene. Vi kan vi definere en differanse mellom devianser på følgende måte, basert på (Dobson 2002, s. 69-80):

$$\Delta D = 2\log \frac{L(\mathbf{b}_1; \mathbf{y})}{L(\mathbf{b}_0; \mathbf{y})} = 2[l(\mathbf{b}_1; \mathbf{y}) - l(\mathbf{b}_0; \mathbf{y})] \quad (7.3)$$

Der:

- $l = \log L$
- \mathbf{b}_0 = maksimumlikelihoodestimatet med p parametre
- \mathbf{b}_1 = maksimumlikelihoodestimatet med q parametre, der $p < q$
- $L(\mathbf{b}_1; \mathbf{y})$ = likelihoodfunksjonen innsatt \mathbf{b}_1 (med q parametre)
- $L(\mathbf{b}_0; \mathbf{y})$ = likelihoodfunksjonen innsatt \mathbf{b}_0 (med $p < q$ parametre)

Her har vi en modell med flere parametere over brøkstreken, og en med færre under brøken. Dersom modellen med færre parametre er betydelig dårligere, blir brøken «stor». Det viser det seg at hvis vi tar logaritmen til dette forholdet og ganger det med 2, så blir dette forholdet tilnærmet $\sim \chi_{q-p}^2$. Dette kan vi vise ved hjelp av en Taylorutvikling.

Vi gjør en Taylorutvikling for $l(\boldsymbol{\beta}; \mathbf{y})$ og får da:

$$l(\boldsymbol{\beta}; \mathbf{y}) = l(\mathbf{b}; \mathbf{y}) + (\boldsymbol{\beta} - \mathbf{b}) \frac{\partial l(\mathbf{b}; \mathbf{y})}{\partial \boldsymbol{\beta}} + \frac{1}{2} (\boldsymbol{\beta} - \mathbf{b})^T \frac{\partial^2 l(\mathbf{b}; \mathbf{y})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} (\boldsymbol{\beta} - \mathbf{b}) + \dots \quad (7.4)$$

Vi antar det som kjent at \mathbf{b} innsatt i den deriverte er $\mathbf{0}$, altså:

$$\frac{\partial l(\mathbf{b}; \mathbf{y})}{\partial \boldsymbol{\beta}} = \mathbf{0} \quad (7.5)$$

Flytter vi første leddet på høyre side over på venstre og ganger med 2 på begge sider får vi da:

$$2[l(\boldsymbol{\beta}; \mathbf{y}) - l(\mathbf{b}; \mathbf{y})] = (\boldsymbol{\beta} - \mathbf{b})^T \frac{\partial^2 l(\mathbf{b}; \mathbf{y})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} (\boldsymbol{\beta} - \mathbf{b}) \quad (7.6)$$

Høyre side kjenner vi igjen som Wald-observatoren (forklares også i (Dobson 2002, s. 69-80)), som er asymptotisk $\sim \chi_p^2$. Vi bruker dette resultatet videre:

$$\begin{aligned} D &= 2[l(\mathbf{b}_{max}; \mathbf{y}) - l(\mathbf{b}; \mathbf{y})] \\ &= 2[l(\mathbf{b}_{max}; \mathbf{y}) - l(\boldsymbol{\beta}_{max}; \mathbf{y})] \\ &\quad - 2[l(\mathbf{b}; \mathbf{y}) - l(\boldsymbol{\beta}; \mathbf{y})] \\ &\quad + 2[l(\boldsymbol{\beta}_{max}; \mathbf{y}) - l(\boldsymbol{\beta}; \mathbf{y})] \end{aligned} \quad (7.7)$$

Her vil første leddet være $\sim \chi_n^2$, (n er antall observasjoner) andre leddet $\sim \chi_p^2$ og siste leddet tilnærmet null dersom modellen med bare p parametre er jevngod med den med maksimalt antall parametre. Dette medfører at D er tilnærmet $\sim \chi_{n-p}^2$.

Denne tilnærmingen er ikke alltid så god, men det viser seg at $D_1 - D_0$ er rimelig godt tilnærmet $\sim \chi_{q-p}^2$, gitt at begge modellene er rimelig bra.

Så vi konkluderer med at da er:

$$\Delta D = D_1 - D_2 = 2 \log \frac{L(\mathbf{b}_1; \mathbf{y})}{L(\mathbf{b}_0; \mathbf{y})} = 2[l(\mathbf{b}_1; \mathbf{y}) - l(\mathbf{b}_0; \mathbf{y})] \sim \chi_{q-p}^2 \quad (7.8)$$

7.2.2 Multinormalfordelingen

På en enkelt laks har vi i prinsippet åtte lengdemålinger og like mange vektmålinger som sammen eller hver for seg kan oppfattes som multinormalfordelte vektorer. Det er derfor naturlig å skrive litt generelt om multinormalfordelte data, og si noe om hvordan man gjør tester på slike data. Deretter returnerer vi igjen til våre konkrete laksedata.

Antar vi at vektoren \mathbf{y} inneholder p observasjoner på samme individ (i våre data er p antall måletidspunkt vi bruker i analysen, eller færre), og at \mathbf{y} er multinormalfordelt $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, får vi følgende sannsynlighetstetthet (Johnson og Wichern 2002):

$$f(\mathbf{y}) = \frac{1}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2}} e^{-(\mathbf{y}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y}-\boldsymbol{\mu})/2} \quad (7.9)$$

I en MANOVA, MANCOVA eller andre modeller skifter vi ut $\boldsymbol{\mu}$ med $\mathbf{X}\boldsymbol{\beta}$ der \mathbf{X} er desingmatrisen og $\boldsymbol{\beta}$ er parametre. For enkelhets skyld bruker jeg $\boldsymbol{\mu}$ i utledningen.

Har vi n uavhengige \mathbf{y}_i er blir simultansannsynlighetstettheten;

$$\begin{aligned} f(\mathbf{y}_1 \dots \mathbf{y}_n) &= \prod_{i=1}^n \left\{ \frac{1}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2}} e^{-(\mathbf{y}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y}_i - \boldsymbol{\mu})/2} \right\} \\ &= \frac{1}{(2\pi)^{np/2} |\boldsymbol{\Sigma}|^{n/2}} e^{-\sum_{i=1}^n (\mathbf{y}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y}_i - \boldsymbol{\mu})/2} \end{aligned} \quad (7.10)$$

Dette gir en likelihood-funksjon som er lik ovenforstående, bare at da er \mathbf{y}_i 'ene gitt og $\boldsymbol{\mu}$ og $\boldsymbol{\Sigma}$ er nå parametre i funksjonen, istedenfor observasjonene $\mathbf{y}_1 \dots \mathbf{y}_n$.

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{np/2} |\boldsymbol{\Sigma}|^{n/2}} e^{-\sum_{i=1}^n (\mathbf{y}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y}_i - \boldsymbol{\mu})/2} \quad (7.11)$$

Vi trenger et teorem fra lineær algebra for å forenkle uttrykket, nemlig følgende;

$$\mathbf{x}^T \mathbf{A} \mathbf{x} = \text{tr}(\mathbf{x}^T \mathbf{A} \mathbf{x}) = \text{tr}(\mathbf{A} \mathbf{x} \mathbf{x}^T) \quad (7.12)$$

Her er x en vektor og A en symmetrisk matrise. Dette er vist i Johnson og Wichern (2002, s. 168-169). Anvender vi dette på eksponenten i likelihoodfunksjonen får vi;

$$\begin{aligned} \sum_{i=1}^n (\mathbf{y}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y}_i - \boldsymbol{\mu}) &= \sum_{i=1}^n \text{tr}[(\mathbf{y}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y}_i - \boldsymbol{\mu})] \\ &= \sum_{i=1}^n \text{tr}[\boldsymbol{\Sigma}^{-1} (\mathbf{y}_i - \boldsymbol{\mu}) (\mathbf{y}_i - \boldsymbol{\mu})^T] \\ &= \text{tr} \left[\boldsymbol{\Sigma}^{-1} \left(\sum_{i=1}^n (\mathbf{y}_i - \boldsymbol{\mu}) (\mathbf{y}_i - \boldsymbol{\mu})^T \right) \right] \end{aligned} \quad (7.13)$$

Setter vi dette inn igjen i eksponenten i likelihoodfunksjonen, får vi:

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{np/2} |\boldsymbol{\Sigma}|^{n/2}} e^{-\text{tr}[\boldsymbol{\Sigma}^{-1} (\sum_{i=1}^n (\mathbf{y}_i - \boldsymbol{\mu}) (\mathbf{y}_i - \boldsymbol{\mu})^T)]/2} \quad (7.14)$$

Vi antar at det er kjent at maksimum likelihood estimatorene for $\boldsymbol{\mu}$ og $\boldsymbol{\Sigma}$ er henholdsvis $\hat{\boldsymbol{\mu}} = \bar{\mathbf{Y}}$ og $\hat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{i=1}^n (\mathbf{Y}_i - \bar{\mathbf{Y}})(\mathbf{Y}_i - \bar{\mathbf{Y}})^T = \frac{n-1}{n} \mathbf{S}$, der $\mathbf{S} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{Y}_i - \bar{\mathbf{Y}})(\mathbf{Y}_i - \bar{\mathbf{Y}})^T$.

Setter vi estimatet $\hat{\boldsymbol{\mu}}$ for $\boldsymbol{\mu}$ inn i likelihoodfunksjonene får vi;

$$L(\hat{\boldsymbol{\mu}}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{np/2} |\boldsymbol{\Sigma}|^{n/2}} e^{-\text{tr}[\boldsymbol{\Sigma}^{-1} (\sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}}) (\mathbf{y}_i - \bar{\mathbf{y}})^T)]/2} \quad (7.15)$$

Videre setter vi inn estimatet for $\boldsymbol{\Sigma} = \hat{\boldsymbol{\Sigma}}$, inn i ligningen, og da får vi bare en konstant igjen oppe i eksponenten;

$$L(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) = \frac{1}{(2\pi)^{np/2}} e^{-np/2} \frac{1}{|\hat{\boldsymbol{\Sigma}}|^{n/2}} \quad (7.16)$$

I ligningen overfor er det bare determinanten som inneholder dataene og parametrene, slik at kun denne blir viktig mht. å vurdere «godheten» til modellen. For å sammenligne godheten til en modell, må den sammenlignes med andre modeller med flere eller færre parametre. Vi ønsker med andre ord å sammenligne simultantettheten til dataene gitt en modell opp mot andre modeller.

Da gjennomføres stegvis en «likelihood-forhold-test» for parvis sammenligning av modellene. Ønsker vi å teste to modeller mot hverandre får vi altså:

$$\begin{aligned}
 D &= 2 \ln \frac{L(\hat{\boldsymbol{\mu}}_0, \hat{\boldsymbol{\Sigma}}_0)}{L(\hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\Sigma}}_1)} = 2 \ln \frac{\frac{1}{(2\pi)^{np/2}} e^{-np/2} \frac{1}{|\hat{\boldsymbol{\Sigma}}_0|^{n/2}}}{\frac{1}{(2\pi)^{np/2}} e^{-np/2} \frac{1}{|\hat{\boldsymbol{\Sigma}}_1|^{n/2}}} \\
 &= 2 \ln \frac{\frac{1}{|\hat{\boldsymbol{\Sigma}}_0|^{n/2}}}{\frac{1}{|\hat{\boldsymbol{\Sigma}}_1|^{n/2}}} \\
 &= 2 \ln \frac{|\hat{\boldsymbol{\Sigma}}_1|^{n/2}}{|\hat{\boldsymbol{\Sigma}}_0|^{n/2}} \\
 &= 2 \ln |\hat{\boldsymbol{\Sigma}}_1|^{n/2} - 2 \ln |\hat{\boldsymbol{\Sigma}}_0|^{n/2} \\
 &= n(\ln |\hat{\boldsymbol{\Sigma}}_1| - \ln |\hat{\boldsymbol{\Sigma}}_0|) \tag{7.17}
 \end{aligned}$$

I Diggle (1990, s.137) ganger han med p som ikke skal være der. Han skriver:

$$D = np(\ln |\hat{\boldsymbol{\Sigma}}_1| - \ln |\hat{\boldsymbol{\Sigma}}_0|) \tag{7.18}$$

Konsekvensen av dette er at differansen blir blåst opp og vi stopper forenklingen av modellen for tidlig i prosessen.

7.2.3 MLE og REML

REML står for «Restricted maximum likelihood», «Residual maximum likelihood» eller «Marginal maximum likelihood» (Searle *et al.* 1992, s. 41). Det kan oppfattes som en litt mer avansert metode enn MLE². I variansanalyse gir REML samme resultater som minste kvadraters metode, fordi frihetsgradene blir tatt hensyn til i beregning av variansen.

MLE har en tendens til å underestimere variansen. Man kan ikke sammenligne modeller med ulike faste effekter ved loglikelihood-forholdstesten når man bruke REML. Man må da bruke MLE, men når man har valgt modell, kan man bruke REML til å finne mer korrekte estimater på variansen (Crawley 2002, s. 296).

I mitt tallmateriale blir estimatene av variansen relativt like ved få parametre, men REML gir alltid litt større variansestimater.

²På norsk brukes sannsynlighetmaksimeringsestimering, og på engelsk «Maximum Likelihood Estimation», derfor MLE.

7.3 ANCOVA

ANCOVA kan defineres som en kombinasjon av variansanalyse og regresjon. Vi betegner ofte generelt forklaringsvariablene som uavhengige størrelser, mens de variablene som skal forklares kalles responsvariabler. Variabelen vi vil ha forklart kalles også den avhengige størrelsen eller responsen.

I en variansanalyse er det kategoriske (nominelle eller diskrete) variabler som er forklaringsvariabler, mens det i en regresjon er kontinuerlige variabler. De kontinuerlige variabler som inngår i ANCOVA betegnes kovariater. Derav navnet ANCOVA (Analysis of Covariance). Ofte vil regresjondelen være av underordnet interesse.

Prinsippielt er dette det samme som multipl regressjon, men jeg foretrekker ANCOVA fordi det er mer i den retning vi bruker metoden. Jeg tenker da at jeg ikke bruker mange mulige forklaringsvariabler. Man er da ofte usikker på om forklaringsvariabler i det hele tatt har en effekt i modellen, slik at såkalte «step-down» og «step-up» prosedyrer er aktuelle.

Jeg vil bruke metoden først og fremst når jeg ser på effekter ett steg fram i tid. Sammenhengen vil være sterkest med forrige måling. Både variansanalysedelen og regresjonsdelen være så komplekse man ønsker, bare man har data nok til å estimere sammenhengene.

Generelt gir en ANCOVA-modell bedre styrke enn en ANOVA-modell, men det forutsetter at det er en god sammenheng mellom kovariaten og responsen. Nedenfor vil jeg si noe om hvordan ANCOVA-modellen brukes når jeg undersøker effekten av lysregime og kjønn i fersksvannsfasen.

7.3.1 Enveis ANCOVA

Vi antar først en ANCOVA-modell for effekten av lysbehandlingen (l) på lengde, vekt eller kondisjonsfaktor på tidspunkt t_2 . Fiskens lengde, vekt eller kondisjonsfaktor har ved t_2 betegnes $Y_{li}(t_2)$ (se nedenfor for forklaring av indekser). Det vil være naturlig å anta at $Y_{li}(t_2)$ avhenger av samme størrelse ved en tidligere måling, kalt $Y_{li}(t_1)$. Vi har da følgende modell:

$$Y_{li}(t_2) = \mu + \alpha_l + \beta[Y_{li}(t_1) - \bar{Y}_{\bullet\bullet}(t_1)] + \epsilon_{lit_2}$$

- $l = 1, 2$ eller 3 , dvs. de ulike lysregimer
- $i = 1, \dots$ antall individer

- $\bullet\bullet$ = gjennomsnitt over indeksene
- μ = forventet lengde, vekt eller kondisjonsfaktor ved t_2
- α_l = effekten av lysbehandlingen
- β = stigningstallet i en lineær regresjon

Størrelsen på β sier noe om hvor mye $Y_{li}(t_2)$ avhenger av $Y_{li}(t_1)$. Sentreringen er gjort for å få meningsfull tolking av parametrene. Variansen er i utgangspunktet lik innen alle behandlingene, men vi skal legge til en variansfunksjon (se nedenfor).

Vi ser at β' er lik for alle behandlingene. Dette er selvfølgelig ikke alltid en realistisk forutsetning. Men vi kan ta hensyn til dette i modellen. Vi lar da β også variere med behandling, og får likningen:

$$Y_{li}(t_2) = \mu + \alpha_l + \beta_l[Y_{li}(t_1) - \bar{Y}_{\bullet\bullet}(t_1)] + \epsilon_{lit_2}$$

Legger vi til en variansfunksjon på feilledet δ_l , kan vi gi rom for at variansen til feilledet kan variere mellom behandlingene. Det gir likningen:

$$Y_{li}(t_2) = \mu + \alpha_l + \beta_l[Y_{li}(t_1) - \bar{Y}_{\bullet\bullet}(t_1)] + \epsilon_{lit_2}\delta_l$$

Dette er et eksempel på en variansfunksjon, som drøftes mer i neste avsnitt.

7.3.2 Variansfunksjoner

Ønsker vi også at modellen skal gi rom for at variansen kan være en fleksibel funksjon, kan vi bruke følgende funksjon:

$$Var(\epsilon_{li}) = \sigma^2 g^2[\mu_{li}, \mathbf{v}_{li}(t_{-1}), \delta_l]$$

Variansen kan altså avhenge av enten forventningen μ_{li} , en kovariat $\mathbf{v}_{li}(t_{-1})$ eller parametre knyttet til gruppe δ_l . Varianter av ovenforstående variansfunksjon tilbys av Splus.

I tabell 7.3.2 på neste side angis variansfunksjoner i lme-rutinen (Crawley 2002) og i nmle-rutinen (Pinheiro og Bates 2000). De samme funksjonene finnes i gls-rutinen i det

Navn	Ligning	Forklaring
varFixed	$Var(\epsilon_{li}) = \sigma^2 v_{li}$	lik i alle grupper,
varIdent	$Var(\epsilon_{li}) = \sigma^2 [\delta_{li}]$	ulik mellom grupper
varPower	$Var(\epsilon_{li}) = \sigma^2 [v_{li}]^{2\gamma}$	avhenger av en potens av kovariaten
varExp	$Var(\epsilon_{li}) = \sigma^2 e^{[2\delta v_{li}]}$	avhenger eksponensielt av kovariaten
varConstPower	$Var(\epsilon_{li}) = \sigma^2 (\delta + v_{li}^{2\gamma})^2$	en konstant + en potens av kovariaten
varComb	$Var(\epsilon_{li}) = \text{produkt}$	av to eller flere funksjoner

Tabell 7.1: Variansfunksjoner i gls-, lme- og nmle-rutinen, der l står for lys og i for individ.

grafiske brukegrensesnittet. Navnene på klasser av variansfunksjonen kan variere mellom ulike rutiner i Splus. Fyller vi inn $y_{li}(t_1)$ istedenfor v_{li} får vi forrige måling som kovariat i tabellen. Dette gir enklere utregninger. Alternativet ville være å estimere forventede verdier til responsen, og så bruke disse som parametre i variansfunksjonen. Med den enklere metoden slipper vi å estimere forventingen først (uten at det rent teknisk regnes som noe problem). To variansfunksjonene jeg finner naturlig å kunne bruke, og å diskutere litt nærmere. Den første er:

$$VarIdent = Var(\epsilon_{li}) = \sigma^2 \delta_{li}^2$$

Der jeg lar variansen variere med lysbehandling. Dette gjør at vi får en veiing av residualene, slik betydningen av residualene blir lik den inverse ($1/\sigma_{li}$) av variansfunksjonen. Se f. eks. (Wooldridge 2003, s. 270–273) om veiet minste kvadraters-estimering i multipl regressjon. Splus bruker MLE ved hjelp av en iterativ prosedyre: Først maksimerer den med hensyn på forventingen, deretter estimerer den variansfunksjon som gir grunnlag for å estimere forventingen på nytt, etc. Ved antatt normalfordelig skulle det gi noenlunde samme resultater som minste kvadraters metode med veiing. Den andre er:

$$VarPower = Var(\epsilon_{li}) = \sigma^2 |v_{li}|^{2\gamma}$$

Denne gir økende varians med økende størrelse på fisken på forrige måling dersom jeg erstatter v_{li} med $y_{li}(t_{-1})$. Dette medfører en nedveing av residualene fra observasjonene til de i utgangspunktet store fiskene, som jeg vil være forsiktig med å bruke.

7.3.3 Toveis ANCOVA

Toveis ANCOVA fungerer i prinsippet som beskrevet ovenfor bare at vi har en additiv effekt av en ny faktor i tillegg til den første. Vi får da følgende modell:

$$y_{li}(t_2) = \mu + \alpha_{(1)l} + \alpha_{(2)k} + \beta[y_{li}(t_1) - \bar{y}_{\bullet\bullet}(t_1)] + \epsilon_{lit_2}$$

Her er de to faktorene representert med $\alpha_{(1)l}$ og $\alpha_{(2)k}$, der l og k står for henholdsvis lys og kjønn.

Ønsker vi å ha med samspill mellom faktorene, kan vi ta det med i form av leddet $\alpha_{(3)lk}$. Skal stigningstallet også variere med en av faktorene, erstatter vi β med β_l eller β_k .

Vi kan også bruke andregradsledd eller tredjegradsledd på kovariaten kombinert med variansfunksjoner. Dette beskrives i neste avsnitt. Vi må dog være varsom så modellen ikke får for mange parametre.

7.3.4 Polynomisk regresjon

Der jeg bruker andre variabler i tillegg til polynomer, blir dette også i realiteten multippel regresjon. Jeg ønsker ikke en kompleks modell, men en enklest mulig modell som er tilpasset dataene.

$$Y_{li}(t_2) = \mu + \alpha_l + \beta_{1l}[y_{li}(t_1) - \bar{y}_{\bullet\bullet}(t_1)] + \beta_2[y_{li}(t_1) - \bar{y}_{\bullet\bullet}(t_1)]^2 + \beta_3[y_{li}(t_1) - \bar{y}_{\bullet\bullet}(t_1)]^3 + \epsilon_{lit_2}\delta_l$$

Her er μ forventet vekt ved andre måling, α_l effekten av lysbehandlingen og β_{1l} stigningstallet i en lineær regresjon. Størrelsen på β_{1l} sier noe om hvor mye $y_{li}(t_2)$ avhenger av $y_{li}(t_1)$. Vi ser at β_{1l} har en indeks for lys, mens β_2 og β_3 er lik for alle lysgruppene. Dette er ikke nødvendigvis en realistisk forutsetning. Men vi antar dette for lettere å kunne sammenligne gruppene.

Det er igjen en variansfunksjon på feilleddet δ_l , fordi vi vil gi rom for at variansen til feilleddet kan variere mellom lysbehandlingene.

Dette er den modellen jeg til slutt fant mest passende til dataene. Det skal vi se nærmere på i kapitlene med resultater m.h.p lys og kjønn.

Her kan $y_{li}(t_1)$ som sagt være lengde, vekt eller kondisjonsfaktor ved tidspunkt (t_1). Lengden på tidspunkt (t_2) vil være best korrelert med lengden på tidspunkt (t_1), vekten på tidspunkt (t_2) med vekten på tidspunkt (t_1) og kondisjonsfaktoren på tidspunkt (t_2) med kondisjonsfaktoren på tidspunkt (t_1). Vi bruker kun samme størrelse på et tidlige tidspunkt, som kovariat.

7.4 MANOVA/MANCOVA

Den biologiske målsettingen i denne oppgaven er å sammenligne effekten av de tre lys-behandlingene og å sammenligne hanner og hunner, samt moden og umoden laks.

Derfor er vi mest interessert i å finne eventuelle forskjeller mellom forventninger til ulike grupper av fisk. Ved enveis ANOVA vil forventningene estimeres gjennom gjennomsnittsverdier for de ulike gruppene av fisk på ulike tidspunkt. Ved toveis ANOVA blir det ikke helt slik, ei heller i en toveis MANOVA. I MANOVA er kovariansmatrisen vesentlig når vi skal gjøre inferens.

Antall observasjoner i hver «celle» bør overstige 20, eller som et minimum må antall observasjoner i hver «celle» overstige antall variabler som inkluderes i undersøkelsen (Hair jr. *et al.* 1998, s. 342).

Et annet forhold som er viktig å være oppmerksom på er at dersom en effekt bare er tilstede på få variabler og videre bare tilstede bare i få behandlingskombinasjoner, vil disse effektene kunne «drukne» blant alle «ikke-effekter» (Johnson og Wichern 2002, s. 332).

Styrken i en MANOVA-test vil også avhenge av korrelasjonen mellom variablene (Cole *et al.* 1994).

7.4.1 Hensikt

En MANOVA-modell kan være anvendelig forutsatt en viss grad av korrelasjon mellom variablene. Hensikten med en MANOVA-modell kan i henhold til (Hair jr. *et al.* 1998, s. 341) være:

- Mutippelt enkeltvariabel problem
 - Hensikten er å unngå problemet med mutippel testing.
 - Ved signifikans går man videre med enkelt-tester.
 - Kalles «protected ANOVA».
- Strukturert multivariabel problem
 - Ser på spesielle kontraster.
 - F.eks. profilanalyse eller repeterte målinger.

- «Ekte» multivariable problem
 - Ser på effekter på hele responsvektoren.

MANOVA-modellen vil i prinsippet kunne ta hensyn til korrelasjonen i dataene, enten den er lik i alle fiskegrupper eller varierer med gruppene.

7.4.2 Modeller

Vi antar først følgende modell, som er en utvidelse toveis-MANOVA til en treveis-MANOVA beskrevet i Johnson og Wichern (2002, s. 309-312). Jeg har her kuttet ut samspill. Hver fisk er representert med en responsvektor i modellen. Dette gir ligningen:

$$y_{i(lkm)} = \mu + \alpha_l + \beta_k + \gamma_m + \epsilon_{i(lkm)} \quad (7.19)$$

som kan skrives mer fullstendig som:

$$\begin{bmatrix} Y_{i(lkm)1} \\ Y_{i(lkm)2} \\ \vdots \\ Y_{i(lkm)8} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_8 \end{bmatrix} + \begin{bmatrix} \alpha_{(l)1} \\ \alpha_{(l)2} \\ \vdots \\ \alpha_{(l)8} \end{bmatrix} + \begin{bmatrix} \beta_{(k)1} \\ \beta_{(k)2} \\ \vdots \\ \beta_{(k)8} \end{bmatrix} + \begin{bmatrix} \gamma_{(m)1} \\ \gamma_{(m)2} \\ \vdots \\ \gamma_{(m)8} \end{bmatrix} + \begin{bmatrix} \epsilon_{i(lkm)1} \\ \epsilon_{i(lkm)2} \\ \vdots \\ \epsilon_{i(lkm)8} \end{bmatrix}$$

Her er μ ene forventningene på de ulike tidspunkt. α ene, β 'ene og γ ene representerer en additiv effekt av henholdsvis lys, kjønn, og modningstatus på hvert tidspunkt. Effektene α , β og γ kan variere med tiden. Det samme gjelder forventningene: μ ene. De har derfor en tidsindeks fra en til åtte. Indeksene l, k og m forklares nedenfor:

- l=lysregime
 - 1 = LDN, 2 = LD12:12 og 3 = LD24:0
- k=kjønn
 - 1 = hanner og 2 = hunner
- m = modningstatus
 - 1 = umoden og 2 = moden

Forenkler vi notasjonen igjen får vi:

$$\begin{bmatrix} y_{i(lkm)1} \\ y_{i(lkm)2} \\ \vdots \\ y_{i(lkm)8} \end{bmatrix} = \mathbf{y}_{i(lkm)} , \quad \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_8 \end{bmatrix} = \boldsymbol{\mu} , \quad \begin{bmatrix} \alpha_{(l)1} \\ \alpha_{(l)2} \\ \vdots \\ \alpha_{(l)8} \end{bmatrix} = \boldsymbol{\alpha}_l , \quad \begin{bmatrix} \beta_{(k)1} \\ \beta_{(k)2} \\ \vdots \\ \beta_{(k)8} \end{bmatrix} = \boldsymbol{\beta}_k$$

$$\begin{bmatrix} \gamma_{(m)1} \\ \gamma_{(m)2} \\ \vdots \\ \gamma_{(m)8} \end{bmatrix} = \boldsymbol{\gamma}_m \quad \text{og} \quad \begin{bmatrix} \epsilon_{i(lkm)1} \\ \epsilon_{i(lkm)2} \\ \vdots \\ \epsilon_{i(lkm)8} \end{bmatrix} = \boldsymbol{\epsilon}_{i(lkm)}$$

Dette gir samme ligning som 7.19:

$$\mathbf{y}_{i(lkm)} = \boldsymbol{\mu} + \boldsymbol{\alpha}_l + \boldsymbol{\beta}_k + \boldsymbol{\gamma}_m + \boldsymbol{\epsilon}_{i(lkm)}$$

som er omtrent samme notasjon som Johnson og Wichern (2002, s. 309-312). Vi gjør følgende antagelser:

$$\sum_{l=1}^3 \boldsymbol{\alpha}_l = \sum_{k=1}^2 \boldsymbol{\beta}_k = \sum_{m=1}^2 \boldsymbol{\gamma}_m = \mathbf{0}$$

Ønsker vi en modell med samspill får vi følgende modell:

$$\mathbf{y}_{i(lkm)} = \boldsymbol{\mu} + \boldsymbol{\alpha}_l + \boldsymbol{\beta}_k + \boldsymbol{\gamma}_m + \boldsymbol{\alpha}\boldsymbol{\beta}_{lk} + \boldsymbol{\alpha}\boldsymbol{\gamma}_{lm} + \boldsymbol{\beta}\boldsymbol{\gamma}_{km} + \boldsymbol{\alpha}\boldsymbol{\beta}\boldsymbol{\gamma}_{lkm} + \boldsymbol{\epsilon}_{i(lkm)}$$

Denne modellen har jeg ikke nok målinger i hver «celle» til å bruke. Vi ser her at modellen er kompleks og at den additive strukturen har brutt sammen. Hver kombinasjon av gruppetilhørighet får en egen forventning. Hensikten med å kjøre en slik modell kan være håp om at samspilleffektene er små og neglisjerbare, slik at vi kan redusere modellen. Samspillet kan betegnes som ordinal eller disordinal (Hair jr. *et al.* 1998, s. 344–345). Ordinal vil si at effekten ikke er lik for mellom nivåene av en faktor gitt en annen faktor, men at retningen er hele tiden den samme. Disordinal innebærer ett tegnskifte for eksempel for kombinasjonen av to faktorer.

Dersom samspillseffektene ikke er neglisjerbare eller vi på et teoretisk grunnlag må regne med et samspill, kan nøsting være et alternativ. Eventuelt kan vi kombinere nøsting og

additivitet. van Belle (2002, s. 143-144) angir at nøsting reduserer antall mulige modeller, og at å ikke velge nøsting fremfor flerveis ANOVA bør ha en sterk begrunnelse. En fullstendig nøsting kan også resultere i en kompleks modell med mange parametre. Tenker vi at modningen har ulik effekt innen kjønnene, mens kjønn gir en additiv effekt i tillegg til lys får vi følgende modell:

$$y_{i(lkm)} = \mu + \alpha_l + \beta_k + \beta_{km} + \epsilon_{i(lkm)}$$

Her representerer β_{km} effekten av modning innenfor ett kjønn. Vi kan ha ulike β_{km} for hver kombinasjon av k og m. Jeg antar at vi kan få følgende restriksjon:

$$\sum_{m=1}^2 \beta_{km}|(k=1) = \sum_{m=1}^2 \beta_{km}|(k=2) = 0$$

men at dette avhenger av kontrasten som velges. Et viktig poeng i forhold til å velge modell er: 1) Ser vi på effekt av modningen på veksten vil en modell med nøstingen: modning innen kjønn kunne være additivt til lys. 2) Betrakter vi derimot endring (bedring) i vekst hos noen fisker som en årsak til modningen vil det være mer naturlig å bare ta med i modellen nøstingen: modning innen kjønn. Det vil si vi ser bort fra lys i modellen fordi det bare er «vekststatusen» som er avgjørende for modningen, ikke forhistorien. Dette er mer i harmoni med en biologisk forklaringmodell hvor man snakker om et beslutningsvindu i en bestemt periode. Altså følgende modell:

$$y_{i(lkm)} = \mu + \beta_k + \beta_{km} + \epsilon_{i(lkm)}$$

7.4.3 MANCOVA

Velger man MANCOVA fremfor MANOVA vil det også få konsekvenser for hvilke hypoteser man i praksis arbeider med. Når man velger en MANCOVA i samme situasjon som forut, så ser vi da på om de modne innen kjønnene er «større» kompensert for startvekten eller smoltvekten eller hva man nå bruker som kovariat. Akkurat som vi tar med lys som en additiv faktor så forandrer det modellen. Vil vi ta med f.eks smoltvekt som kovariat ser vi mer på veksten i sjøen enn størrelsen, som igjen er mer tråd med Thorpe (1994).

Vi skal og være klar over at vi da forutsetter mer enn i en MANOVA. Det skal i utgangspunktet være en lineær sammenheng mellom respons og kovariat. Denne skal være lik i alle grupper. Man kan i prinsippet slakke på kravene og gjøre mer komplekse tilpassinger. En MANCOVA modell med startvekt som kovariat og modning nøstet innen kjønn blir:

$$\begin{bmatrix} Y_{i(lkm)2} \\ Y_{i(lkm)3} \\ \vdots \\ Y_{i(lkm)8} \end{bmatrix} = \begin{bmatrix} \mu_2 \\ \mu_3 \\ \vdots \\ \mu_8 \end{bmatrix} + \begin{bmatrix} \beta_{(k)2} \\ \beta_{(k)3} \\ \vdots \\ \beta_{(k)8} \end{bmatrix} + \begin{bmatrix} \beta_{(km)2} \\ \beta_{(km)3} \\ \vdots \\ \beta_{(km)8} \end{bmatrix} + \begin{bmatrix} \beta_{(2)r}(x_{1i} - x_1) \\ \beta_{(3)r}(x_{1i} - x_1) \\ \vdots \\ \beta_{(8)r}(x_{1i} - x_1) \end{bmatrix} + \begin{bmatrix} \epsilon_{i(lkm)2} \\ \epsilon_{i(lkm)3} \\ \vdots \\ \epsilon_{i(lkm)8} \end{bmatrix}$$

Her har $Y_{i(lkm)1}$ blitt en forklaringsvariabel $(x_{1i} - x_1)$ som er sentret, mens leddet $\beta_{(2)r}(x_{1i} - x_1)$ er regresjonen fra kovariaten inn på $Y_{i(lkm)2}$. De to indeksene på $\beta_{(2)r}$ står for henholdsvis hvilken variable som forklares ved regresjonen og r for regresjon.

I valget mellom MANOVA og MANCOVA, blir spørsmålet relatert til denne oppgaven om vi har data som passer til MANCOVA og hvilke forandringer en MANCOVA eventuelt gjør med hypotesene.

7.4.4 Kovariansmatrise

Uansett modell vil vektoren med feilledd antas uavhengig mellom hver enkelt fisk, men at det eksisterer en korrelasjon mellom de enkelte feilledd innen hver fisk. Kovariansmatrisen til feilleddene antas å være lik for hver gruppe av fisk, eller for sammenslåtte grupper av fisk, for eksempel for alle fisk innen hver lysgruppe. Eventuelt er den lik for all fisk. Det siste gjør modellen enklest.

Feilleddene antas å være multinormalfordelt $\sim N_p(\mathbf{0}, \Sigma)$, der Σ er kovariansmatrisen og p er antall målinger, som vil være lik åtte eller færre, avhengig av hvor mange måletidspunkt vi vil se på.

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{bmatrix}$$

I utgangspunktet antas kovariansmatrisen til feilleddene å være ustrukturert, dvs. den estimeres bare direkte fra feilleddene. Da får vi:

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n (\mathbf{y}_i - \hat{\mathbf{y}}_i)(\mathbf{y}_i - \hat{\mathbf{y}}_i)^T = \frac{n-1}{n} \mathbf{S} = \frac{n-1}{n} \begin{bmatrix} s_{11} & s_{12} & \dots & s_{1p} \\ s_{21} & s_{22} & \dots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \dots & s_{pp} \end{bmatrix}$$

Som i realitet bare er empiriske varianser og kovarianser basert på n istedenfor $n - 1$. Men det er verd å merke seg at modellen påvirker kovariansmatrisen fordi det er $\hat{\mathbf{Y}}$ som inngår i estimeringen av feilleddene.

For å redusere antall parametre kan vi også legge på en struktur på kovariansmatrisen. Det diskuteres i neste delkapittel.

7.4.5 Testobservatorer

Aktuelle testobservatorer for MANOVA er Pillais spor, Wilks lambda, Hotelling-lawley og Roys største rot (alle defineres nedenfor). Asymptotisk skal de gi like resultater. Som regel gir disse testene like resultater med rimelig mange data, men de er følsomme for ulike typer avvik. For eksempel antydes det at Pillais spor er mindre følsom for ikke normalitet og ubalanserte data. Roy største rot ser ut til å ha størst styrke når en gruppe på eventuelt bare én variabel viser avvik fra de andre gruppene. Jeg fokuserer i utgangpunktet på Wilks lambda fordi den tilsvarer en loglikelihoodforholdstesten, og Roys største rot pga. følsomhet for effekter i få «celler» samt Pillais spor p.g.a robustheten.

Testobservatorene defineres på følgende måte i en enveis MANOVA:

- Wilks lambda = $\frac{|W|}{|B+W|}$
- Hotelling-lawley spor = $tr[\mathbf{B}\mathbf{W}^{-1}]$
- Pillai spor = $tr[\mathbf{B}(\mathbf{B} + \mathbf{W})^{-1}]$
- Roys største rot = maximum eigenverdi av $\mathbf{B}(\mathbf{B} + \mathbf{W})^{-1}$

der matrisene \mathbf{W} og \mathbf{B} som tilsvarer SS_{error} og $SS_{behandling}$ i en enveis ANOVA. \mathbf{W} og \mathbf{B} defineres på følgende måte:

$$\mathbf{W} = \sum_{l=1}^g \sum_{i=1}^{n_l} (\mathbf{y}_{li} - \bar{\mathbf{y}}_l)(\mathbf{y}_{li} - \bar{\mathbf{y}}_l)^T$$

$$\mathbf{B} = \sum_{l=1}^g n_l (\bar{\mathbf{y}}_l - \bar{\mathbf{y}})(\bar{\mathbf{y}}_l - \bar{\mathbf{y}})^T$$

der g , b og n_l er henholdsvis antall grupper (lysregimer) og antall observasjoner i hver gruppe (lysregime).

For en toveis MANOVA (balansert) med samspill, slik som beskrevet i Johnson og Wichern (2002, s. 210–213) blir matrisene (der g , b og n er henholdsvis antall lysregimer, antall kjønn og et fast antall observasjoner i hver «celle») som følger:

$$SSP_{lys} = \sum_{l=1}^g bn(\bar{\mathbf{y}}_{l\bullet} - \bar{\mathbf{y}}_{\bullet\bullet})(\bar{\mathbf{y}}_{l\bullet} - \bar{\mathbf{y}}_{\bullet\bullet})^T$$

$$SSP_{kjønn} = \sum_{k=1}^b gn(\bar{\mathbf{y}}_{\bullet k} - \bar{\mathbf{y}}_{\bullet\bullet})(\bar{\mathbf{y}}_{\bullet k} - \bar{\mathbf{y}}_{\bullet\bullet})^T$$

$$SSP_{samspill} = \sum_{l=1}^g \sum_{k=1}^b n(\bar{\mathbf{y}}_{l k \bullet} - \bar{\mathbf{y}}_{l\bullet\bullet} - \bar{\mathbf{y}}_{\bullet k \bullet} + \bar{\mathbf{y}}_{\bullet\bullet\bullet})(\bar{\mathbf{y}}_{l k \bullet} - \bar{\mathbf{y}}_{l\bullet\bullet} - \bar{\mathbf{y}}_{\bullet k \bullet} + \bar{\mathbf{y}}_{\bullet\bullet\bullet})^T$$

$$SSP_{error} = \sum_{l=1}^g \sum_{k=1}^b \sum_{i=1}^n (\mathbf{y}_{lki} - \bar{\mathbf{y}}_{l k \bullet})(\mathbf{y}_{lki} - \bar{\mathbf{y}}_{l k \bullet})^T$$

$$SSP_{Total} = \sum_{l=1}^g \sum_{k=1}^b \sum_{i=1}^n (\mathbf{y}_{lki} - \bar{\mathbf{y}}_{\bullet\bullet\bullet})(\mathbf{y}_{lki} - \bar{\mathbf{y}}_{\bullet\bullet\bullet})^T$$

Poenget med å vise dette er at vi ser ved ubalanserte data så er ikke n lik i alle grupper, slik at det er mindre trivielt hvordan matrisen beregnes. Searle *et al.* (1992, s. 168) skriver at Hendersons metoder for ubalansert ANOVA vil bli erstattet av MLE og andre teknikker.

Vi ser også at SSP_{error} tilsvarer estimert kovariansmatrise, men ikke delt på antall frihetsgrader. I mine data antar jeg at SSP_{error} og SSP_{total} ikke er så forskjellig, fordi vi trolig ikke får forklart så mye av «variansen» med modellen. Dette er ikke fordi modellen er dårlig men fordi det rett slett er mye individuell variasjon.

7.4.6 En ujustert R^2

For å kunne sammenligne modeller var det meningen å regne ut en slags ujustert R^2 basert på følgende formel etter forslag fra Hans Karlsen:

$$R^2 = 1 - \frac{|\hat{\Sigma}_1|}{|\hat{\Sigma}_0|}$$

Imidlertid kom jeg for sent i gang med dette, slik at det utgår. Generelt for at R^2 skal gi mening må modellene være spesialtilfeller av hverandre og basert på samme metode (Anderson-Sprecher 1994). Se også (Kvalseth 1985).

7.4.7 Ikke-parametriske metoder

Det finnes ikke-parametriske metoder også for MANOVA. for eksempel som beskrevet i Anderson (2001). En ulempe generelt knyttet til ikke parametriske metoder er at vi ikke alltid får parameter estimater. I en del tilfeller vil det da ikke være mulig å lage konfidensintervaller (van Belle 2002, s. 119–120). Det vil si at vi ikke får svar på for eksempel hvor mye bedre vokser en moden fisk kontra en umoden. På den annen side kan det være gunstig å bruke ikke-parametriske metoder, som en ekstra kontroll av de parametriske metodene. Resultane av begge metoder bør være konsistente.

7.4.8 Foreløpig vurdering

Jeg hadde i utgangspunktet mest tro på en MANOVA-tilnærming. Dette fordi vekstkurver lett kan gi mistilpasninger. Mistilpasningene kommer av at fiskene rett og slett ikke alltid vokser slik funksjonene angir. Se for eksempel Troynikov (1998). Men vekstkurver har minst en fordel slik jeg ser det, siden de i prinsippet kan takle manglende målinger. Hviken metode som er «best» vil blant annet avgjøres av antall målinger per individ og eventuelt antall parametre i vekstkurven. Vi går derfor over til å se på en tilnærming med vekstkurver ved hjelp lineære blandede modeller.

7.5 Lineære blandede modeller (LME)

Bruken av denne metoden er vist i et eget appendix. Lineære blandede modeller, heretter kalt LME, inneholder både «faste» og «varians» komponenter. Med det menes at med hensyn på noen forklaringsvariabler er vi bare interessert i variansen, mens for andre vil vi estimere en fast effekt. Vi kan også si (noe forenklet) at en fast effekt tilsvarer om et parameter gjeldene for hele populasjonen, mens varianskomponenter har med individuelle eksperimentelle enheter å gjøre.

Lineære blandede modeller behandles i S-plus i rutinen «lme» som står for «Linear Mixed-Effects» på engelsk. En annen aktuell rutine for vekstdata er «nlme» som står for

«Nonlinear Mixed-Effects» (Pinheiro og Bates 2000). Generelt kan blandede lineære modeller uttrykkes på formen:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i \quad (7.20)$$

Her står $\mathbf{X}_i\boldsymbol{\beta}$ for den faste delen av modellen, tilsvarende en ordinære lineær modell. Varianskomponenten er $\mathbf{Z}_i\mathbf{b}_i$ som er knyttet til individuelle effekter. Disse (\mathbf{b}_i) oppfattes som stokastiske og kalles derfor på engelsk «random effects», men jeg vil kalle dem individ effekter, fordi det avspeiler måten de blir brukt på her.

Typisk vil man anta at \mathbf{b}_i er fordelt $N_m(\mathbf{0}, \boldsymbol{\Psi})$. Der indeksen m står for antallet individ effekter. Som regel vil \mathbf{Z}_i være en del av designmatrisen. Videre vil vi anta at $\boldsymbol{\epsilon}_i$ er fordelt $N_p(\mathbf{0}, \sigma^2\mathbf{I})$, men det er mulig å anta heteroskedastisitet og ulike former for autokorrelasjon.

7.5.1 Modeller

Hvordan en slik modell fungerer illustreres best gjennom anvendelse. Vi vil her se på hvordan modellen kan anvendes på våre data. Konkret i vår situasjon tilpasser vi enten en rett linje eller en tredjegradsfunksjon til hver enkelt fisk. Vi ser først på en modell der en rett linje er tilpasset fire av målingene til hver enkelt fisk. Det innebærer en individ effekt på både skjæringspunktet og stigningstall. Ser vi for enkelhet skyld vekk fra lys, kjønn og modning, vil vi uttrykke modellen på følgende måte³;

$$\begin{bmatrix} Y_{i3} \\ Y_{i4} \\ Y_{i5} \\ Y_{i6} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 100 \\ 1 & 200 \\ 1 & 300 \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \beta_0 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 100 \\ 1 & 200 \\ 1 & 300 \end{bmatrix} \begin{bmatrix} b_{1i} \\ b_{2i} \end{bmatrix} + \begin{bmatrix} \epsilon_{i3} \\ \epsilon_{i4} \\ \epsilon_{i5} \\ \epsilon_{i6} \end{bmatrix}$$

Notasjonen er hentet fra (Jones 1993). Her har jeg latt tiden gå fra 0 til 300 i hundredagers steg. Det som skiller denne modellen fra «vanlig regresjon» er leddene b_{1i} og b_{2i} som er random effekter. Disse gir rom for en regresjonslinje tilpasset hver enkelt fisk. Vi er interessert i variansen til disse effektene, slik at vi kan estimere usikkerheten til de faste effektene. Vi ender altså opp med to typer feilledd. Dette er da ett spesialtilfelle av den generelle formen for blandede lineære modeller. Ønsker vi å skrive modellen uten bruk av lineær algebra, får vi:

³Notasjonen er hentet fra Jones (1993) men jeg bruker b for individeffekt.

$$y_i(t) = (\alpha_0 + b_{1i}) + (\beta_0 + b_{2i})t + \epsilon_i \quad (7.21)$$

Legger vi på en effekt av lys, på skjæringspunktet i modellen, får vi følgende for lysgruppe nr. 2.:

$$\begin{bmatrix} Y_{i3} \\ Y_{i4} \\ Y_{i5} \\ Y_{i6} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 100 & 1 & 0 \\ 1 & 200 & 1 & 0 \\ 1 & 300 & 1 & 0 \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \beta_0 \\ \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 100 \\ 1 & 200 \\ 1 & 300 \end{bmatrix} \begin{bmatrix} b_{1i} \\ b_{2i} \end{bmatrix} + \begin{bmatrix} \epsilon_{i3} \\ \epsilon_{i4} \\ \epsilon_{i5} \\ \epsilon_{i6} \end{bmatrix}$$

Der α_2 er ett tillegg for lysgruppe nr. 2 relativt til gruppe nr 1. Og for lysgruppe nr 3. får vi:

$$\begin{bmatrix} Y_{i3} \\ Y_{i4} \\ Y_{i5} \\ Y_{i6} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 100 & 0 & 1 \\ 1 & 200 & 0 & 1 \\ 1 & 300 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \beta_2 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 100 \\ 1 & 200 \\ 1 & 300 \end{bmatrix} \begin{bmatrix} b_{1i} \\ b_{2i} \end{bmatrix} + \begin{bmatrix} \epsilon_{i3} \\ \epsilon_{i4} \\ \epsilon_{i5} \\ \epsilon_{i6} \end{bmatrix}$$

Der α_3 er ett tillegg for lysgruppe nr. 3 relativt til gruppe nr. 1.

Ønsker vi videre at lys skal kunne ha en effekt på stigningstallet i tillegg til skjæringspunktet, får vi følgende modell for lysgruppe 2:

$$\begin{bmatrix} Y_{i3} \\ Y_{i4} \\ Y_{i5} \\ Y_{i6} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 100 & 1 & 0 & 100 & 0 \\ 1 & 200 & 1 & 0 & 200 & 0 \\ 1 & 300 & 1 & 0 & 300 & 0 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \beta_1 \\ \alpha_2 \\ \alpha_3 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 100 \\ 1 & 200 \\ 1 & 300 \end{bmatrix} \begin{bmatrix} b_{1i} \\ b_{2i} \end{bmatrix} + \begin{bmatrix} \epsilon_{i3} \\ \epsilon_{i4} \\ \epsilon_{i5} \\ \epsilon_{i6} \end{bmatrix}$$

Her vil α_2 være effekten lysbehandling nr. to, reativt til nr. 1 på skjæringspunktet. Likeledes vil β_2 være effekten lysbehandling nr. to, reativt til nr. en på stigningstallet.

$$\begin{bmatrix} Y_{i3} \\ Y_{i4} \\ Y_{i5} \\ Y_{i6} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 100 & 0 & 1 & 0 & 100 \\ 1 & 200 & 0 & 1 & 0 & 200 \\ 1 & 300 & 0 & 1 & 0 & 300 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \beta_1 \\ \alpha_2 \\ \alpha_3 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 100 \\ 1 & 200 \\ 1 & 300 \end{bmatrix} \begin{bmatrix} b_{1i} \\ b_{2i} \end{bmatrix} + \begin{bmatrix} \epsilon_{i3} \\ \epsilon_{i4} \\ \epsilon_{i5} \\ \epsilon_{i6} \end{bmatrix}$$

Uten lineær algebra gir dette gir da følgende likning for en rett linje, dersom vi har en additiv effekt av lys både på intersepten og stigningstallet.

$$y_i(t) = (\alpha_0 + \alpha_1 X_2 + \alpha_2 X_3 + b_{1i}) + (\beta_0 + \beta_1 X_2 + \beta_2 X_3 + b_{2i})t + \epsilon_i \quad (7.22)$$

Her er X_2 og X_3 indikatorvariabler, som er 0 eller 1 avhengig om hvilken lysgruppe vi har (null i begge dersom lysgruppe=1, avhengig av kontrast). Videre er α_2 og α_3 additive effekter av henholdsvis lys på intersepten. Likeså er β_2 og β_3 additive effekter av henholdsvis lys på stigningstallet.

Tilsvarende modifikasjoner av modellen får vi når vi legger på effekt av kjønn og modning. Faktisk får vi bare to ekstra parameter for kjønn og det samme fra modning. Dette siden kjønn og modning bare har to kategorier. Dermed legges for eksempel parameteren for hannene inn i α_1 og β_1 men det legges til en ekstra α og β som blir parametrene for hunnene. Tilsvarende blir det for umoden og moden. Parameteren for umoden legges inn i α_1 og β_1 mens det legges til en ekstra α og β som blir parametrene for moden.

Men vi kan ta med en illustrasjon av hvordan en modell med et tredjegrads polynom fungerer. Her har vi tillatt individ effekter på alle leddene, slik at vi får fire individ effekter.

$$\begin{bmatrix} Y_{i3} \\ Y_{i4} \\ Y_{i5} \\ Y_{i6} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 100 & 100^2 & 100^3 \\ 1 & 200 & 200^2 & 200^3 \\ 1 & 300 & 300^2 & 300^3 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 100 & 100^2 & 100^3 \\ 1 & 200 & 200^2 & 200^3 \\ 1 & 300 & 300^2 & 300^3 \end{bmatrix} \begin{bmatrix} b_{1i} \\ b_{2i} \\ b_{3i} \\ b_{4i} \end{bmatrix} + \begin{bmatrix} \epsilon_{i3} \\ \epsilon_{i4} \\ \epsilon_{i5} \\ \epsilon_{i6} \end{bmatrix}$$

Denne modellen har jeg bestemt meg for å bruke (som nevnt i 1.2), i samband med kjønn og modningsproblematikk. Da kombinert med nøsting og valgte kontraster. (Modellene beskrives fullt ut i resultatdelene.)

7.5.2 Korrelasjonstrukturer

Splus tilbyr en del korrelasjonsstrukturer på feilledet (Pinheiro og Bates 2000). Disse er gjengitt i tabell 7.5.2 etter Crawley (2002).

Navn	Forklaring
Tidsserie	
corAR1	autoregressiv prosess (AR1)
corARMA	autoregressiv prosess (ARMA)
corBand	«bånd» korrelasjon struktur
corCAR1	kontinuerlig autoregressiv prosess (kontinuerlig AR1)
CorComsymm	konstant korrelasjon
Romlige	
CorExp	eksponensiell romlig korrelasjon
corGaus	gaussisk romlig korrelasjon
corLin	lineær romlig korrelasjon
corRatio	ratio kvadratisk romlig korrelasjon
corSpher	sfærisk romlig korrelasjon
corStrat	stratifisert korrelasjon struktur
corSymm	generalisert: ingen struktur

Tabell 7.2: Korrelasjonstrukturer i lme- og nmle-rutinen.

Aktuelle strukturer for meg er corAR1, corCAR1 eller CorExp. Som nevnt under delkappitlet om longitudinelle data, mener jeg at ingen av disse er godt nok tilpasset vårt datasett.

Splus tilbyr en del strukturer på kovariansmatrisen til «random effektene» (Pinheiro og Bates 2000, s. 158). Disse er gjengitt i tabell 7.5.2.

Navn	Forklaring
pdBlocked	blokk diagonal
pdCompSymm	sammensatt-symmetri struktur
pdDiag	diagonal
pdIdent	multippel av en identitet
pdSymm	generel positiv-definit matrise

Tabell 7.3: Strukturer kovariansmatrisen til random effektene i lme- og nmle-rutinen.

Her har jeg valgt å bruke pdSymm helt konsekvent, altså ingen spesiell struktur.

7.5.3 Likelihoodfunksjonen

Lar vi for enkelthets skyld feilleddene være uavhengige og identisk fordelt. Modellen kan da re-uttrykkes:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\epsilon}^* \quad (7.23)$$

Der $\boldsymbol{\epsilon}^* = \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i$. Siden $\boldsymbol{\epsilon}_i^*$ er en sum av to uavhengige multinormalfordelte vektorer med forventning lik $\mathbf{0}$ og kovarians-varians matrise lik $\sigma^2\boldsymbol{\Sigma}_i$ kan vi skrive:

$$p(\mathbf{y}_i|\boldsymbol{\beta}, \boldsymbol{\phi}, \sigma^2) = \frac{1}{(2\pi)^{1/2}|\boldsymbol{\Sigma}|^{1/2}} e^{(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}) / 2\sigma^2} \quad (7.24)$$

Der $\boldsymbol{\phi}$ er parametre knyttet til kovarians-varians matrisen til individeffektene $\boldsymbol{\Psi}$ og $\boldsymbol{\Sigma}_i = \mathbf{I} + \mathbf{Z}_i\boldsymbol{\Psi}\mathbf{Z}_i^T / \sigma^2$. For gitt verdi av $\boldsymbol{\theta}$ som representerer parametre knyttet til $\boldsymbol{\Psi}$, kan vi maksimere likelihoodfunksjon m.h.p. $\boldsymbol{\beta}$ og σ^2 . Vi får da:

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \left[\sum_{i=1}^n \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \right]^{-1} \left[\sum_{i=1}^n \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{y}_i \right] \quad (7.25)$$

$$\hat{\sigma}^2(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}(\boldsymbol{\theta}))^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}(\boldsymbol{\theta})) \quad (7.26)$$

I følge (Pinheiro og Bates 2000, s. 66) er ligning (7.25) og ligning (7.26) ikke lette å løse beregningsmessig ved progammering. De angir derfor en annen metode som prinsipielt skal løse dette. Jeg beskriver ikke i detalj hvordan dette gjøres, men gir bare en liten antydning:

Generelt ved lineære blandede modeller får vi en likelihood-funksjon som er lik følgende:

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2 | \mathbf{y}) = p(\mathbf{y} | \boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2) \quad (7.27)$$

- $\mathbf{y} = \mathbf{y}_1 \cdots \mathbf{y}_n$
- $\boldsymbol{\beta}$ = parametre knyttet til faste effekter
- $\boldsymbol{\theta}$ = parametre knyttet til kovariansmatrisen til random effektene
- σ^2 = variansen til feilleddene

For enkelhets skyld lar vi igjen feilleddene uavhengige og identisk fordelt. Som vanlig er y_i 'ene gitt og β , θ og σ er nå parametre i funksjonen, istedenfor observasjonene $y_1 \cdots y_n$.

Bruker vi loven om total sannsynlighet og definisjonen på betinget sannsynlighet, vil det være mulig å splitte opp uttrykket. Vi integrere altså først med hensyn på b_i 'ene, deretter splittes uttrykket ved å betinge på b_i .

$$\begin{aligned} L(\beta, \theta, \sigma^2 | \mathbf{y}) &= \prod_{i=1}^n p(y_i | \beta, \theta, \sigma^2) \\ &= \prod_{i=1}^n \int p(y_i, \mathbf{b}_i | \beta, \theta, \sigma^2) d\mathbf{b}_i \\ &= \prod_{i=1}^n \int p(y_i | \mathbf{b}_i, \beta, \sigma^2) p(\mathbf{b}_i | \theta, \sigma^2) d\mathbf{b}_i \end{aligned} \quad (7.28)$$

Dette innebærer en veiing av observasjonene over en sannsynlighetsfordeling for individeffektene. Vi antar videre at den betingede fordelingen til y_i gitt \mathbf{b}_i , er multinormal og at \mathbf{b}_i er multinormalfordelt. Da får vi:

$$p(y_i | \mathbf{b}_i, \beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{p/2}} e^{-(y_i - \mathbf{X}_i\beta - \mathbf{Z}_i\mathbf{b}_i)^T (y_i - \mathbf{X}_i\beta - \mathbf{Z}_i\mathbf{b}_i) / 2\sigma^2} \quad (7.29)$$

$$p(\mathbf{b}_i | \theta, \sigma^2) = \frac{1}{(2\pi)^{q/2} |\Psi|^{1/2}} e^{-\mathbf{b}_i^T \Psi^{-1} \mathbf{b}_i / 2} \quad (7.30)$$

der q er antall individeffekter (lengden av vektoren \mathbf{b}_i). Setter vi disse to sannsynlighetstetthetene inn i likningen overfor får vi:

$$\prod_{i=1}^n \int \frac{1}{(2\pi\sigma^2)^{p/2}} e^{-(y_i - \mathbf{X}_i\beta - \mathbf{Z}_i\mathbf{b}_i)^T (y_i - \mathbf{X}_i\beta - \mathbf{Z}_i\mathbf{b}_i) / 2\sigma^2} \frac{1}{(2\pi)^{q/2} |\Psi|^{1/2}} e^{-\mathbf{b}_i^T \Psi^{-1} \mathbf{b}_i / 2} d\mathbf{b}_i \quad (7.31)$$

Rydder vi litt får vi:

$$\prod_{i=1}^n \left[\frac{1}{(2\pi\sigma^2)^{p/2}} \frac{1}{(2\pi)^{q/2} |\Psi|^{1/2}} \cdot \int \exp \left(-[(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)^T (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i) + (\sigma^2 \mathbf{b}_i^T \Psi^{-1} \mathbf{b}_i)] / 2\sigma^2 \right) d\mathbf{b}_i \right] \quad (7.32)$$

Jeg forstår dette som at de bruker en veiing av dataene over en gitt sannsynlighetsfordeling av «random effektene». De bruker videre en pseudo-data representasjon og orthogonal-triangular dekomponering. Den interesserte leser henvises til (Pinheiro og Bates 2000, s. 58-96).

7.5.4 LmList-rutinen

Dersom vi hadde brukt LmList-rutinen (Pinheiro og Bates 2000) med et tredjegrads polynom tilpasset 4 datapunkter ville vi fått perfekt tilpasning og ingen residualer. Denne rutinen ville brukt minste kvadraters metode på hver fisk, slik at koeffisientene etterpå måtte betraktes som en transformasjon av datene. Lme-rutinen fungerer altså anderledes og gir oss residualer i dette tilfellet. Det har å gjøre med at Lme-rutinen optimaliserer med hensyn på ett sett av kurver som har en fordeling rundt en forventet kurve.

7.6 Diggle's modell

En måte å håndtere vekstforsøk på er å betrakte dem som longitudinelle data, (han bruker riktig nok betegelsen repeterte målinger) slik det er beskrevet i (Diggle 1990, s. 135-164). Han inkorporer individeffekter og autokorrelasjon. Som nevnt tidligere i 7.14 kan man også velge mellom en MANOVA eller en vekstkurvetilnærming. Jeg har valgt å bruke en ordinær MANOVA og Lineære blandede lineære modeller som løsning. Vi skal derfor se litt på Diggle's modell, slik at vi ser at de metodene jeg har brukt fanger de samme problemstillingene.

Det er viktig å presisere at den bruk av Diggles modeller jeg ser på her ikke er anvendt på de data jeg nå behandler. Men ved bruk av ny måleteknologi, vil man kunne ha mange flere målinger på hver fisk, slik kurvene som da tilpasses er mer realistiske. Det medfører at residualene blir mer realistiske, og at hans modell er svært aktuell. Diggle's modell tar også grep på problemstillingen med ulike måletidspunkt.

7.6.1 Modellen

Vekstforløpet til hver fisk er her, en stokastisk variabel; $Y_i(t)$ fordelt $N_p(\boldsymbol{\mu}_i(t), \boldsymbol{\Sigma})$. Han bruker følgende likning til å beskrive veksten;

$$Y_i(t) = \boldsymbol{\mu}_i(t) + \mathbf{Z}_i(t) \quad (7.33)$$

Her er $\boldsymbol{\mu}_i(t)$ en forventningsvektor som er lik for alle fiskene tilhørende samme gruppe eller behandling. $\mathbf{Z}_i(t)$ er et «feilledd» som er uavhengig av forventningsvektoren.

Forskjellen mellom MANOVA og Diggles modell er at forventningsleddet og feilleddet kan modelleres mer fleksibelt hos ham. Forventningen kan om man vil bestå av en vekstkurve, mens feilleddet kan inneholde en stasjonær autoregressiv prosess (Se (Brockwell og Davis 2002, s. 15) for definisjon).

7.6.2 Forventningen

Forventningen $\boldsymbol{\mu}_i(t)$ kan også modelleres slik at hver gruppe eller behandling av fisken følger en kurve. I sin enkleste form blir det at $\mu_i(t) = \beta_1 + \beta_2 t$, $\mu_i(t) = \beta_1 + \beta_2 t + \beta_3 t^2$ eller $\mu_i(t) = \beta_1 + \beta_2 t + \beta_3 t + \beta_4 t^3$. Eventuelt kan man bruke en av modellene angitt i avsnittet om vekstmodeller. Alternativt kan man estimere forventningen ved hjelp av MANOVA (eller MANCOVA metoder).

Ved bruk av modell som angitt ved ligning (7.33) vil man først estimere forventningene, og etterpå estimere kovariansmatrisen ut fra residualene i en iterativ prosedyre (MLE eller REML).

I tillegg kan vi ha individeffekter, som omtales i neste avsnitt.

7.6.3 «Feilleddet»

Feilleddet $Z_i(t_j)$ (restleddet) ved et gitt tidspunkt t_j , kan bestå av uavhengige målefeil; V_{ij} , eventuelt en individfaktor; U_i og et ledd fra en autoregressiv svakt stasjonær prosess; $S_i(t_j)$. Se (Brockwell og Davis 2002, s. 15) for definisjon. Da får vi:

$$Z_i(t_j) = U_i + V_{ij} + S_i(t_j) \quad (7.34)$$

Siden våre laksedata har en varians som er økende med tiden, vil de ikke tilfredsstillende kravet i modellen om at $S_i(t_j)$ er en svakt stasjonær prosess.

Man kan da transformere (log eller annen passende transformasjon) dataene slik at de passer bedre inn i Diggles modell. Hensikten med transformasjonen er å stabilisere variansen, som vil ha en generell tendens til å øke med tiden eller veksten.

Et annet alternativ er å se på SGR, slik som Sandland og McGilchrist (1979) argumenter for er mer passende for en autoregressiv stasjonær prosess. Dette sporet har jeg ikke fulgt videre, blant annet fordi SGR ikke var den eneste størrelsen jeg var interessert i.

Struktur og antall parametre i kovariansmatrisen Σ vil avhenge av tilnærming. Antall parametre i en ustrukturert kovariansmatrise er $\frac{1}{2}p(p+1)$ der p er antall målinger pr. individ.

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{bmatrix}$$

Vi kan anta like tidsintervaller mellom målingene og at dataene er transformert slik at vi har lik varians. Da får vi at σ_{jk} bare avhenger av antall tidsintervaller mellom målingene. Da blir antall parametre redusert til 2; en for varians og en for kovariansmatrisen σ_{jk} , der j og k er etterfølgende tidspunkt. Vi kan også bruke følgende betegnelser: $\sigma_{jk} = \gamma(|t_j - t_k|) = \gamma(u)$, der u er antall tidsintervall $|t_j - t_k| = (u)$.

Det er ikke like tidsintervaller mellom målingene i våre data. Derfor antar vi heller at σ_{jk} bare avhenger av lengden til tidsintervallet mellom målingene, får vi $\sigma_{jk} = \gamma(|t_j - t_k|) = \gamma(u)$. Der tidsintervallet $|t_j - t_k| = (u)$. Da blir antall parametre redusert til p dersom alle intervallene har ulike lengde og vi har p tidsintervaller.

Bruker vi en parametrisert modell på $\gamma(u)$ kan vi redusere antall parametre til en ved å anta $\rho(u) = e^{-\alpha u}$. Dette tilsvarer en AR[1] prosess, der korrelasjonen minker «jevnt» med tidsavstanden mellom målingene. Gitt ligning (7.34) på forrige side, får vi følgende varians og kovarians for $Z_i(t_j)$;

$$\text{var}[Z_i(t_j)] = \gamma(0) = v^2 + \tau^2 + \sigma^2$$

$$\begin{aligned}
\text{cov}[Z_i(t_j), Z_i(t_k)] &= \gamma(u) = E[Z_i(t_j)Z_i(t_k)] \\
&= E[(U_i + V_{ij} + S_i(t_j))(U_i + V_{ik} + S_i(t_k))] \\
&= E[U_i^2] + E[S_i(t_j)S_i(t_k)] \\
&= v^2 + \sigma^2\rho(|t_j - t_k|)
\end{aligned} \tag{7.35}$$

Det innebærer at kovariansmatrisen avhenger av individeffektene og variansen til den autoregressive svakt stasjonære prosessen og av hvordan $\rho(|t_j - t_k|)$ utvikler seg med tiden. Det er da å mulig skrive $\Sigma = \sigma^2 V$ der V inneholder strukturen til kovariansmatrisen. Eller om vi vil $\Sigma(\boldsymbol{\phi}) = \sigma^2 V(\boldsymbol{\phi})$ der $\boldsymbol{\phi}$ er en vektor som inneholder parametrene til kovariansmatrisen.

Vi skal se litt mer på strukturerte kovariansmatriser seinere, men først ser vi på likelihoodfunksjonen.

7.6.4 Likelihoodfunksjonen

Antar vi multinormalfordeling og lik kovariansmatrise til alle individ, får vi følgende likelihoodfunksjon:

$$L(\boldsymbol{\theta}, \boldsymbol{\phi}, \sigma^2) = \frac{1}{(2\pi)^{np/2} |\Sigma(\boldsymbol{\phi})|^{n/2}} e^{-\sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\theta})^T \{\Sigma(\boldsymbol{\phi})\}^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\theta}) / 2} \tag{7.36}$$

Der $\boldsymbol{\phi}$ nå er parametre knyttet til å parametrisere kovariansmatrisen. Videre vil $\mathbf{X}_i \boldsymbol{\theta}$ være estimerte verdier knyttet til forventningene, uansett om vi velger en MANOVA eller vekstkurvetilnærming.

Trekker vi σ^2 ut fra kovariansmatrisen og skriver $\Sigma = \sigma^2 V$ der V inneholder strukturen til kovariansmatrisen. Vi får da en loglikelihood-funksjon lik:

$$l(\boldsymbol{\theta}, \boldsymbol{\phi}, \sigma^2) = -\frac{1}{2} \left[n \log \sigma^2 + n \log |V(\boldsymbol{\phi})| + \sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\theta})^T \{V(\boldsymbol{\phi})\}^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\theta}) / \sigma^2 \right] \tag{7.37}$$

Gitt verdier for $\boldsymbol{\phi}$ blir sannsynlighetsmaksimeringsestimatoren for $\boldsymbol{\theta}$ lik:

$$\hat{\boldsymbol{\theta}}(\boldsymbol{\phi}) = \left[\sum_{i=1}^n \mathbf{X}_i^T \{V(\boldsymbol{\phi})\}^{-1} \mathbf{X}_i \right]^{-1} \left[\sum_{i=1}^n \mathbf{X}_i^T \{V(\boldsymbol{\phi})\}^{-1} \mathbf{y}_i \right] \tag{7.38}$$

Det er her verdt å merke seg at kovariansmatrisen påvirker estimatet av θ gjennom en «veiting». Sannsynlighetsmaksimeringsestimatoren $\hat{\sigma}^2$ for σ^2 blir:

$$\hat{\sigma}^2(\boldsymbol{\phi}) = \frac{1}{n} \sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\theta}}(\boldsymbol{\phi}))^T \{\mathbf{V}(\boldsymbol{\phi})\}^{-1} (\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\theta}}(\boldsymbol{\phi})) \quad (7.39)$$

Bruker vi uttrykkene i ligning (7.39) og ligning (7.38) på forrige side og setter inn i ligning (7.37) på forrige side får vi i henhold til Diggle (1990, s. 153):

$$l(\boldsymbol{\phi}) = -\frac{1}{2} [n \log \sigma^2 + n \log |\mathbf{V}(\boldsymbol{\phi})|] \quad (7.40)$$

Dette blir:

$$l(\boldsymbol{\phi}) = -\frac{1}{2} [\log |\hat{\sigma}^2 \mathbf{V}(\boldsymbol{\phi})|^n] \quad (7.41)$$

Som kun involverer $\boldsymbol{\phi}$. Etter å ha funnet $\hat{\boldsymbol{\phi}}$ er maksimum likelihood estimatene for θ og σ^2 henholdsvis $\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\phi}})$ og $\hat{\sigma}^2(\hat{\boldsymbol{\phi}})$. Et hovedpoeng ved Diggles modell er vi kan ha individuell ulike kovariansmatriser. Dette er nødvendig dersom ulike fisk er målt på ulike tidspunkt. Da bytter vi bare ut $\mathbf{V}(\boldsymbol{\phi})$ med $\mathbf{V}_i(\boldsymbol{\phi})$ i formlene overfor. Det innebærer parametrisering av kovariansmatrisen.

7.6.5 Strukturerte kovariansmatriser

Som vist nevnt tidligere er $p(p+1)/2$ antall parametre som må estimeres i en ustrukturert kovariansmatrise når antall variabler er lik p .

Fordelen ved strukturerte kovariansmatriser er at vi får færre antall parametre å estimere. Dette gir flere frihetsgrader og en strukturert kovariansmatrise kan og gi sikrere estimat av kovariansene, fordi vi bruker hele datamaterialet til å bestemme noen få parametre. Forutsetningen er imidlertid at strukturen må stemme rimelig bra med virkeligheten.

Den mest realistiske strukturen knyttet til «våre» data, er lettest å observere på korrelasjonsmatrisen. Der vil korrelasjonen minke med økende tidsintervall.

Dette tilsier for eksempel den eksponensielle korrelasjonsfunksjonen: $\rho(t) = e^{-\alpha(t)}$, eller den Gaussisk korrelasjonsfunksjonen: $\rho(t) = e^{-\alpha(t^2)}$. Den eksponensielle korrelasjonsfunksjonen tilsvarer en AR(1) modell.

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{bmatrix}$$

At variansen er økende med tiden passer ikke med en AR(1) modell. Vanligvis vil man da transformere slik at variansen blir mer stabil eller noenlunde lik. Dette anbefales i Diggle (1990).

Et moment som kan være viktig fordi vi har målt fisken over to år, da kan sesongavhengig vekst påvirke hvor godt $\rho(t) = e^{-\alpha(t)}$ parametriserer kovariansmatriseren.

På grunn av at fisken vokser svært lite i vintersesongen så blir t ikke den riktige variabel til å parametriserer kovariansmatriseren. Her har jeg et par alternative løsninger. De jeg noe om i neste avsnitt.

7.6.6 Mitt løsnings forslag

Det er viktig å presisere at den bruk av Diggles modeller jeg foreslår ikke er tenkt anvendt på de data jeg nå behandler. Men ved bruk av ny måleteknologi, vil man kunne ha mange flere målinger på hver fisk, slik kurvene som da tilpasses er mer realistiske. Det medfører at residualene blir mer realistiske. I en slik situasjon mener jeg at den «forbedringen» som jeg forslår kan ha noe for seg. Spesielt også ved bruk av MANOVA, vil en slik struktur fremtre, men da må en ha like måletidspunkt.

Mulige alternativer vil være å bruke $(t_0 - t_g) \sum_{i=1}^g (d_i - \bar{d})$ der $(d_i - \bar{d})$ representerer avvik i daglengden, eller SGR^t der SGR er gjennomsnittlig daglig spesifikk tilvekst. Altså tar vi med en sesongkomponent eller noe som avspeiler realisert vekst i perioden. Dette er en videreutvikling av Diggles modell, men den kan begrunnes ut fra teori i kapitèlet om vekstkurver og direkte ut i fra spredningsplott av mine data.

En tredje alternativ måte å håndtere dette på, et etter min oppfatning at man baserer seg på både at korrelasjonen minker slik som ved en eksponensielle korrelasjonsfunksjon og at variansen øker med tiden. Man får da to parametre som gir kovariansmatriseren.

Slik er ofte den reelle strukturen til typiske biologiske vekstdata().

Dersom man studere prosesser og utvikle modeller som fører til slike kovariansmatriser, vil dette ikke nødvendigvis gi mer «riktige» vekstmodeller generelt, siden man empirisk vet allerede at strukturen ofte er slik.

7.6.7 Variogram

Variogram kan brukes til å bestemme formen på kovariansmatrisen. Dersom tidsintervallene ikke er like mellom alle målinger, kan det være gunstig å bruke variogrammet. Det er imidlertid mest nyttig når ulike fisker er målt på ulike tidspunkt.

Variogrammet er et plott der vi har avstand mellom målingene, dvs. i vårt tilfelle lengden av tidsintervallene, k på horisontal akse og det estimerte variogram $\hat{V}(k)$ på den vertikale aksene.

$V(k)$ defineres på følgende måte;

$$\begin{aligned}
 V(k) &= \frac{1}{2}E[(Y(t) - Y(t-k))^2] \\
 &= \frac{1}{2}E[Y(t)^2 - 2Y(t)Y(t-k) + Y(t-k)^2] \\
 &= \frac{1}{2}[\gamma(0) - 2\gamma(k) + \gamma(0)] \\
 &= \frac{1}{2}[2\gamma(0) - 2\gamma(k)] \\
 &= \gamma(0) - \gamma(k) \\
 &= \gamma(0) - \gamma(k) \frac{\gamma(0)}{\gamma(0)} \\
 &= \gamma(0) \left[1 - \frac{\gamma(k)}{\gamma(0)}\right] \\
 &= \gamma(0) [1 - \rho(k)]
 \end{aligned} \tag{7.42}$$

Det medfører at $\hat{V}(k)$ beregnes utfra $\hat{\gamma}(0) - \hat{\gamma}(k)$. Vi får da startverdier til å optimalisere likelihood-funksjonen.

7.6.8 Foreløpig konklusjon

I Diggle's modell får individeffektene og autokorrelasjonen kun konsekvenser for kovariansmatrisen.

Siden jeg har såpass rikelig med data for de fleste grupper, lar kovariansmatrisen seg estimere rimelig bra uten parametrisering. Antar jeg videre lik kovariansmatrise for alle grupper, vil i enda sterkere grad behovet for å parametrisere kovariansmatrisen bortfalle. Konklusjonen blir derfor at behovet for å parametrisere kovariansmatrisen ikke er kritisk i mitt datasett.

Altså: Siden AR[1] prosessen og individeffektene kun påvirker kovariansmatrisen kan jeg like godt bruke en vanlig MANOVA. I MANOVA blir kovariansmatrisen estimert, og gitt multinormalfordeling er kovariansmatrisen sammen med μ eller $X\beta$ sufficient.

Når jeg bruker vekstkurvetilnærmingen via lme modulen i Splus blir individeffektene ivaretatt. I denne modulen kan jeg også modellere inn autokorrelasjon.

Dersom målingene ikke var på samme tidspunkt for hvert individ, ville det likevel være aktuelt å parametrisere kovariansmatrisen. Da vil vi kunne bruke variogram til hjelp med å bestemme parametrene.

7.6.9 Lme-rutinen og «Diggles modell»

Får vi så brukt Diggle modell ved å bruke lme-rutinen eller nlme-rutinen? Svaret mitt er ja, hvis vi ser på modellen slik den er beskrevet i (Diggle 1990, s. 134-164). Dette vil jeg begrunne på ved følgende sammenligninger:

Vekstforløpet til hver fisk er hos Diggle, en stokastisk variabel; $Y_i(t)$ fordelt $N_p(\mu_i(t), \Sigma)$. Følgende likning beskriver veksten;

$$Y_i(t) = \mu_i(t) + Z_i(t) \quad (7.43)$$

Her er $\mu_i(t)$ en forventningsvektor som er lik for alle fiskene tilhørende samme gruppe eller behandling. $Z_i(t)$ er et «feilledd» som er uavhengig av forventningsvektoren.

Feilleddet $Z_i(t_j)$ (restleddet) ved et gitt tidspunkt t_j , består av uavhengige målefeil; V_{ij} , eventuelt en individfaktor; U_i og et ledd fra en autoregressiv stasjonær prosess; $S_i(t_j)$. Da får vi:

$$Z_i(t_j) = U_i + V_{ij} + S_i(t_j) \quad (7.44)$$

Dette er de samme elementene som tilbys av lme-rutinen, gjennom å bruke en individ-faktor etter å ha transformert dataene, og så bruke corAR1 eller CorExp på residualene, slik som nevnt i tabell 7.5.2.

Dette ser også andre ut til å mene (Crawley 2002, s. 681). Men hvorvidt man kan ha individuelle kovariansmatriser, som i andre sammenhenger er et hovedpoeng i (Diggle 1990, s. 134-164), har jeg ikke prøvd ut, eller sett nærmere på.

8 Effekt av lysbehandling

Problemstilling

I dette kapitlet ser jeg på effekten av lysbehandlingene i ferskvannsfasen har på vekt, lengde og kondisjonsfaktor i mai, altså andre måling.

Effekten av lys på vekstmønster er behandlet grundig av andre. Derfor ser jeg ikke på dette. Istedenfor ser jeg først og fremst etter effekter av lysbehandlingene som kan ha betydning for analyse av forskjeller mellom kjønnene og forskjeller mellom modne og umodne laks. Vekstforskjeller mellom moden og umoden fisk er den biologiske hovedproblemstillingen i min tilnærming til datasettet.

Derimot bruker jeg lys som en ekstra forklaringsvariabel når jeg studerer eventuelle kjønnsforskjeller (neste kapittel). Jeg kommer også inn på ferskvannsfasen der. Første del av sjøvannsfasen (måling 3–6) analyseres der i forbindelse med kjønn som problemstilling.

Metoder

Ved bruke av ANCOVA-modeller, får man kompensert for effekten av individuelle startvekter eller startlengder,

For kondisjonsfaktor er en ANCOVA-modell ikke egnet, fordi det er liten sammenheng mellom kondisjonsfaktorene på de to aktuelle tidspunktene. I starten av forsøket antar jeg at vekt, lengde og kondisjonsfaktor er lik for alle lysgruppene (slik vi har sett i kapittel 6). I forbindelse med kondisjonsfaktor ser jeg på muligheten til å bruke andre mål enn den tradisjonelle Fultons kondisjonsfaktor.

8.1 Ferskvannsfasen

Notasjon

I ferskvannsfasen har vi to måletidspunkt per individ, henholdsvis 15. september 1987 og 26. mai 1988. Første måling av vekt, lengde og kondisjonsfaktor for individ i og lysbehandling l betegnes henholdsvis:

- $w_{li}(t_1)$, $l_{li}(t_1)$ og $k_{li}(t_1)$.

Andre måling betegnes tilsvarende:

- $w_{li}(t_2)$, $l_{li}(t_2)$ og $k_{li}(t_2)$.

De tre aktuelle lysbehandlingene er:

- 1 = SNL, 2 = SNL+B og 3 = LD24:0.

Kontraster

I de to første modellene bruker jeg kontrastene cbind («contr.treatment», «contr.poly»). Da blir koeffisientene i utskriften som angitt nedenfor. I de to første modellene representerer koeffisientene i hovedsak estimerte forskjeller mellom lysgruppene.

- Intercept = $\mu + \alpha_1$
- lys2 = $\alpha_2 - \alpha_1$
- lys3 = $\alpha_3 - \alpha_1$
- vekt- eller lengdesentrert1 = β_{11}
- I(vekt- eller lengdesentrert1²) = β_2
- I(vekt- eller lengdesentrert1³) = β_3
- lys2vekt- eller lengdesentrert1 = $\beta_{12} - \beta_{11}$
- lys3vekt- eller lengdesentrert1 = $\beta_{13} - \beta_{11}$

Der «vekt- eller lengdesentrert1» er variablen fratrukket gjennomsnittet. De enkelte koeffisientene tilsvarende de i modellene for vekt og lengde (gitt nedenfor i ligning (8.1) på neste side og ligning (8.2) på side 125). Tilsvarende er også forsøkt angitt i tabell 8.1 på side 124.

8.1.1 Vekt

Modell

Jeg har forsøkt å tilpasse ulike modeller for å finne en god og adekvat modell. Etter litt prøving og feiling, der jeg gradvis bygget ut modellen, har jeg bestemt meg for en modell med tredjegradspolynom (gitt nedenfor).

Jeg vil gi rom for at variansen til feilledet kan variere mellom lysbehandlingene. Dette gjør at vi får en veiing av residulene, slik at betydningen av residualene blir ulik for hver lysgruppe. Residualene veies med den inverse av variansfunksjonen.

$$W_{li}(t_2) = \mu + \alpha_l + \beta_{1l}[W_{li}(t_1) - \bar{W}(t_1)] + \beta_2[W_{li}(t_1) - \bar{W}(t_1)]^2 + \beta_3[W_{li}(t_1) - \bar{W}(t_1)]^3 + \epsilon_{lit_2}\delta_l \quad (8.1)$$

Der μ er forventet vekt ved andre måling, α_l er effekten av lysbehandlingene og β_{1l} er stigningstallet i en lineær regresjon. Størrelsen på β_{1l} sier noe om hvor mye $w_{li}(t_2)$ avhenger av $w_{li}(t_1)$.

Vi ser at β_{1l} har en indeks for lys, mens β_2 og β_3 er lik for alle lysgruppene. Dette er ikke nødvendigvis en realistisk forutsetning. Men vi antar dette for lettere å kunne sammenligne gruppene.

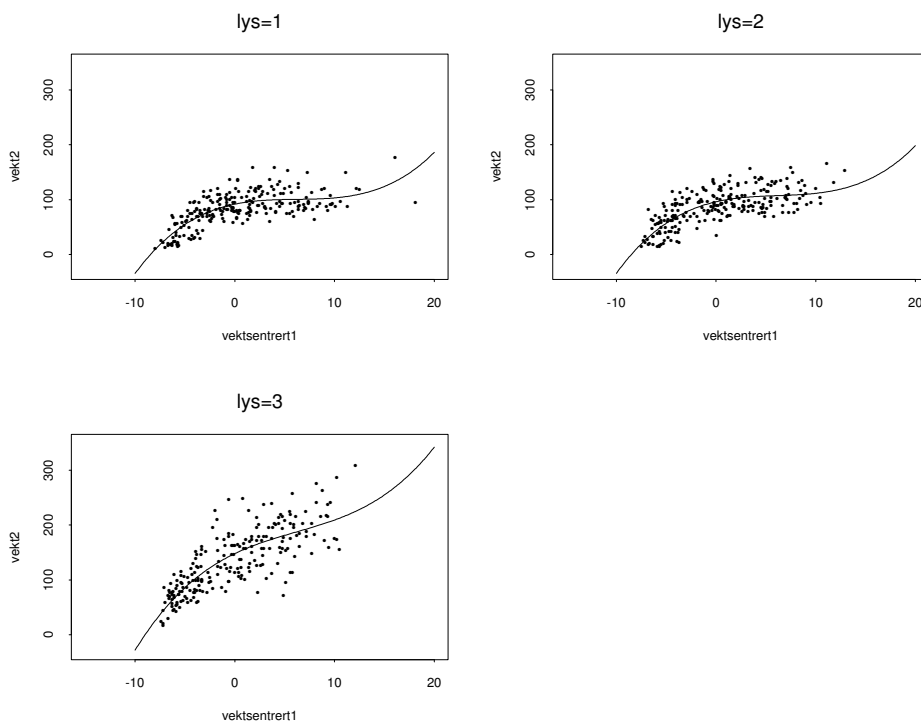
Resultater

Ved tilpasning av denne modellen fikk jeg i gls-rutinen i Splus resultater som vist i tabell 8.1 på side 124 og utskrift som vist i appendiks A. Estimert modell er vist i figur 8.1 på neste side og diagnostiske plot er vist i figur A.1. Tabell 8.2 på side 124 viser 95-prosent konfidensintervall for parametrene i modellen. Bredden på konfidensintervallene sier noe om hvor sikre estimatene er. Variansfunksjonen er «varIdent», dvs. ulik varians mellom lysbehandlingene.

Hovedresultatet som vist i tabell 8.1 på side 124 og tabell 8.2 på side 124, er at LD24:0 gruppen har et løft på vel 50 gram, (51-62 i 95-prosent konfidensintervall) i forhold til SNL målt sentralt i fordelingen av vekt1. Dette er en stor effekt. Lysgruppen SNL+B estimeres til å ha ett løft på ca. 4 gram (0,5-8 i 95-prosent konfidensintervall) relativt til SNL. Dette er mindre enn en tiendedel av effekten til LD24:0-gruppen. Effekten er relativt liten, og upålitelig.

Tabell 8.1 på neste side viser at alle effektene unntatt lys2vektsentert1, er statistisk signifikante med $\alpha = 0.05$. Videre er resten av p-verdiene unntatt lys2 på mindre enn en tiendedels promille. Dette forteller oss at de fleste effektene sannsynligvis ikke er «støy» i modellen.

I alle gruppene ser det ut til å være en sammenheng mellom vekten ved start av forsøket og smoltvekten (vist i figur 8.1). For de to gruppene med SNL, ser det ut til å være en utflating. Når startvekten er over en viss størrelse gir den ikke utslag i større smoltvekt. I lysgruppe LD24:0 ser det ikke ut til å være en slik utflating. Derimot estimeres stignings-tallet for denne gruppen til å være mer enn dobbelt så stort som stigningstallet for de to andre gruppene. Det er β_{11} som skifter vinkelen til modellen mellom lysgruppe. (β_{11} , β_2 og β_3 kan imidlertid ikke sees isolert fra hverandre.)



Figur 8.1: Spredningsplott av vekten 26. mai 1988 (vekt2) mot vekten (minus middelvekten til alle grupper) 15. september 1987 (vektsentert1). Linjene representerer modellen. Plottet er delt opp i tre, etter lysbehandling. 1 = SNL, 2 = SNL+B og 3 = LD24:0.

Vi ser at variansen omkring modellen ser ut til å være mellom 50 og 100 prosent større for LD24:0 gruppen. Dette vil jeg komme tilbake til når jeg også ser på kjønn som en faktor. I

Tabell 8.1: Koeffisienter for vekt2: tredjegradspolynom.

ancova1v	del av modell	verdi	st.feil	t-verdi	p-verdi
(Intercept)	$\mu + \alpha_1$	91.7	1.6	55.6	< .0001
lys2	$\alpha_2 - \alpha_1$	4.3	1.96	2.2	0.0277
lys3	$\alpha_3 - \alpha_1$	56.1	2.73	20.6	< .0001
vektsentrert1	β_{11}	3.8	0.37	10.16	< .0001
(vektsentrert1) i andre	β_2	-0.6	0.05	-11.27	< .0001
(vektsentrert1) i tredje	β_3	0.03	0.0046	6.65	< .0001
lys2vektsentrert1	$\beta_{12} - \beta_{11}$	0.4	0.41	0.98	0.3278
lys3vektsentrert1	$\beta_{13} - \beta_{11}$	4.97	0.57	8.78	< .0001

Tabell 8.2: Konfidensintervall (95 prosent) for effekter, varians og residual. Modell: tredjegradspolynom for vekt2.

intervallv	Nedre	Estimat	Øvre
Effekter			
(Intercept)	88.5	91.7	95.0
lys2	0.476	4.33	8.18
lys3	50.78	56.2	61.50
vektsentrert1	3.047	3.78	4.51
I(vektsentrert1 i andre)	-0.675	-0.57	-0.47
I(vektsentrert1) i tredje	0.0219	0.0310	0.0402
lys2vektsentrert1	-0.404	0.402	1.21
lys3vektsentrert1	3.86	4.97	6.09
Variansfunksjon			
lys2	0.93	1.05	1.19
lys3	1.48	1.68	1.92
st. feil Residual			
st. feil	18.96	20.75	22.72

første omgang ser det ut til at LD24:0-gruppen er mer varierende i veksten enn de andre gruppene. Dette så vi også av histogrammene i figur 6.9 på side 67 og de påfølgende figurene i kapittel 6.

8.1.2 Lengde

Modell

Jeg antar lik modell for lengde som for vekt i ferskvannsfasen (tredjegradpolynom med ulikt førstegradsledd og ulik varians mellom lysgruppene):

$$L_{li}(t_2) = \mu + \alpha_l + \beta_1[L_{li}(t_1) - \bar{L}(t_1)] + \beta_2[L_{li}(t_1) - \bar{L}(t_1)]^2 + \beta_3[L_{li}(t_1) - \bar{L}(t_1)]^3 + \epsilon_{lit_2}\delta_l. \quad (8.2)$$

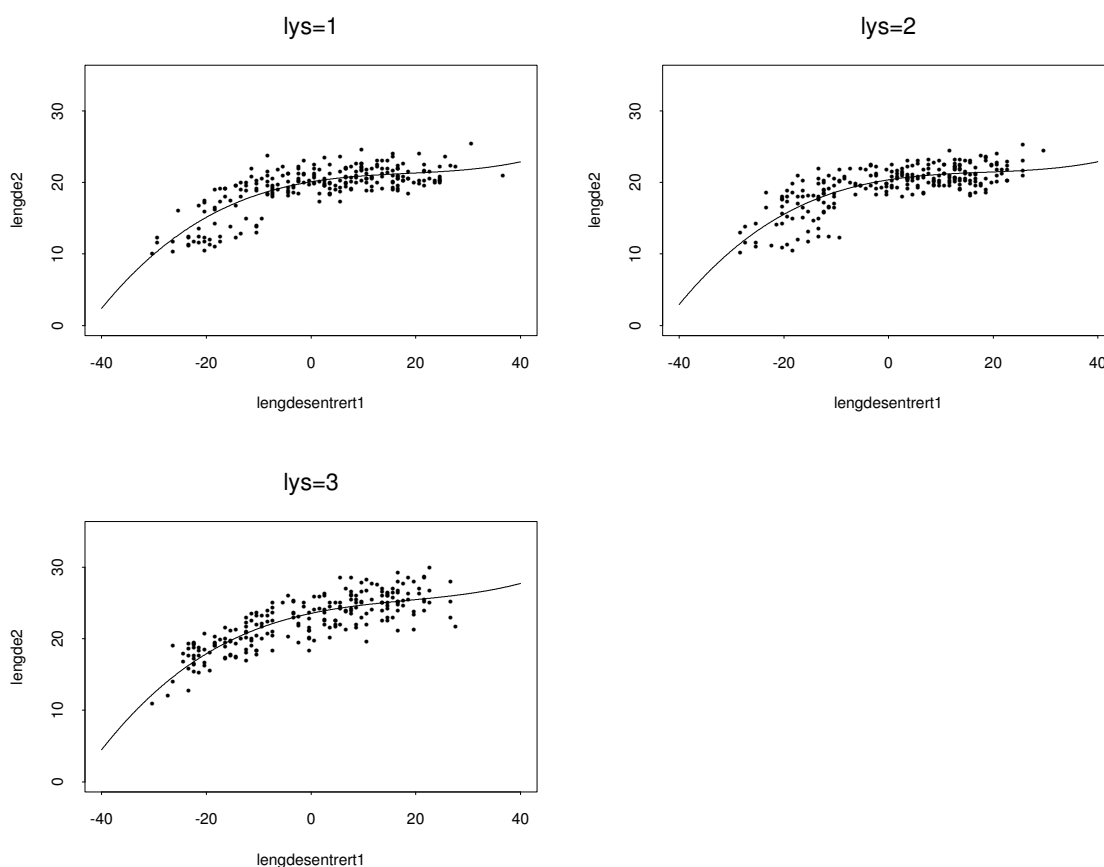
Resultater

Ved kjøring av modellen i gls-rutinen har jeg fått følgende resultater angitt i tabell 8.3 på side 127 og tabell 8.4 på side 127. Hvordan modellen går gjennom dataene er vist i figur 8.2 på neste side. Diagnostiske plot er vist i figur A.2. Disse viser avik fra normalfordeling i halene, samt at det er to grupper av fisk innen lysregimene med naturlig lys.

Lengdevekst er generelt en mer stabil prosess enn vektutvikling. Fisk vil vanligvis vokse i lengde også når den sultes. Vi gjør hovedsaklig samme observasjoner for lengde som for vekt. Lengden til LD24:0-gruppen estimeres til å være 3–4 cm mer enn SNL målt sentralt i fordelingen av lengde1 (vist i tabell 8.3 på side 127 og tabell 8.4 på side 127). Samme fenomen med utflating hos de to gruppene med SNL observeres med hensyn på lengde (vist i figur 8.2 på neste side).

Jeg har ikke andre kommentarer enn at mønsteret er det samme som for vekt i samme fase. Dersom vi speiler datene ned på y-aksen ser vi at at bimodaliteten er mer utviklet av figur 8.2 på neste side. Variansene er ikke så ulike. Det kommer av at de fiskene som «henger igjen» i lengdeutvikling i begge gruppene med SNL bidrar sterkt til variansen i disse gruppene.

Ser vi på lengde i forhold til vekt vil mer dramatiske forhold fremtre (se neste avsnitt).



Figur 8.2: Spredningsplott av lengden 26. mai 1988 (lengde2) mot startlengde (minus middelstartlengden til alle grupper) Linjene representerer modellen. Plottet er delt opp i tre, etter lysbehandling. 1 = SNL, 2 = SNL+B og 3 = LD24:0.

Tabell 8.3: Koeffisienter for lengde2: tredjegradspolynom.

ancova1L	verdi	st.feil	t-verdi	p-verdi
(Intercept)	20.2	0.15	139.3	< .0001
lys2	0.22	0.17	1.28	0.2
lys3	3.4	0.19	17.72	< .0001
len1cmsentrert	0.98	0.13	7.5	< .0001
I(len1cmsentrert) i andre	-0.47	0.039	-12.1	< .0001
I(len1cmsentrert) i tredje	0.11	0.023	4.66	< .0001
lys2len1cmsentrert	0.076	0.12	0.61	0.54
lys3len1cmsentrert	0.49	0.13	3.65	0.0003

Tabell 8.4: Konfidensintervall (95 prosent) for effekter, varians og residual. Modell: tredjegradspolynom for lengde2.

intervall1L	Nedre	Estimat	Øvre
Effekter			
(Intercept)	19.96	20.2	20.5
lys2	-0.12	0.22	0.56
lys3	2.98	3.4	3.7
len1cmsentrert	0.72	0.98	1.24
I(len1cmsentrert)i andre	-0.55	-0.47	-0.39
I(len1cmsentrert) i tredje	0.06	0.11	0.15
lys2len1cmsentrert	-0.17	0.08	0.32
lys3len1cmsentrert	0.23	0.49	0.76
Varians funksjon			
lys2	0.84	0.96	1.10
lys3	0.99	1.14	1.31
St. feil residual			
St. feil	1.61	1.77	1.95

8.1.3 Kondisjonsfaktor

Modeller

Vi så av figur 6.7 at det ikke er en «pen» lineær sammenheng mellom kondisjonsfaktorene på de to aktuelle tidspunktene. Dette gjør en ANCOVA-modell tilsvarende den vi brukte for lengde og vekt lite egnet. Jeg viser likevel hvordan ligningen ville vært:

$$K_{li}(t_2) = \mu + \alpha_l + \beta_{1l}[K_{li}(t_1) - \bar{K}(t_1)] + \beta_{2l}[K_{li}(t_1) - \bar{K}(t_1)]^2 + \epsilon_{lit_2}\delta_l. \quad (8.3)$$

En enklere modell for kondisjonsfaktor vil være:

$$K_{li}(t_2) = \mu + \alpha_l + \epsilon_{li}. \quad (8.4)$$

Dette tilsvarer en vanlig enveis ANOVA. Men vi har svært ulik varians og metoden gir lavere styrke enn metoden nedenfor.

En alternativ tilnærming er å bruke en ANCOVA-modell på sammenheng mellom lengde og vekt på samme tidspunkt. Vi undersøker da om fisk av samme lengde har ulik vekt i de tre lysregimene. Vi kan da bruke følgende modell:

$$W_{li(t_2)} = \mu + \alpha_l + \beta_{1l}[L_{li}(t_1)]^3 + \epsilon_{lit_2}\delta_l. \quad (8.5)$$

Ulempen med denne modellen er at vi ikke får skjæringpunkt i null. Å låse modellen til tredje potens, er gjort for lettere kunne sammenligne med den tradisjonelle kondisjonsfaktoren. En mer realistisk modell er fleksibel på potensen til lengden (også mellom gruppene). Da vil tillegget avhenge av lengden, noe som er mer realistisk. Vi får da følgende modell:

$$W_{li(t_2)} = \alpha_l[L_{li}(t_1)]^{\beta_l}\epsilon_{lit_2}\delta_l. \quad (8.6)$$

Tar vi den naturlige logaritmen på begge sider får vi:

$$\ln W_{li(t_2)} = \ln[\alpha_l[L_{li}(t_1)]^{\beta_l}\epsilon_{lit_2}\delta_l] \quad (8.7)$$

som igjen gir:

$$\ln W_{li(t_2)} = \ln \alpha_l + \beta_l \ln[L_{li}(t_1)] + \ln[\epsilon_{lit_2}\delta_l] \quad (8.8)$$

Da kan jeg bruke en vanlig lineær regresjon gitt at feilledet er normalfordelt. Det forutsetter egentlig at ϵ_{lit_2} var lognormalfordelt, slik at bare én av de gitte modeller kan ha normalfordeling. I den første modellen er feilledet additiv og i den neste et produkt før

Tabell 8.5: Koeffisienter: vekt forklart ved lengde i tredje + konstantledd.

kondis1	verdi	st.feil	t-verdi	p-verdi
(Intercept)	3.50	1.01	3.46	0.0006
lys2	0.68	0.69	1.0	0.3183
lys3	7.34	2.88	2.6	0.0109
$I(len2^3)$	0.0102	0.000114	90.2	< .0001

transformering. Dette er også inkonsistent. Hovedpoenget med disse modellene er å vise at variansene er ulike, og at dette kan estimeres ved en variansfunksjon. Meningen er også å vise at spredningsplott av lengde mot vekt lettere får fram ulikheten i «variens» enn tradisjonelle histogram av kondisjonsfaktor.

Resultater

Resultatene fra modellen fastlåst i tredje potens er angitt i tabell 8.5 og tabell 8.6 på neste side.

Det er også verd å merke seg at LD24:0-gruppen generelt estimeres til å være litt tyngre (mellom 1,7–13) for lik lengde. Jeg tror imidlertid ikke på et fast løft slik modellen angir. Jeg har mer tro på den avvikende større variansen på mellom 4,6 og 6 hos LD24:0 gruppen gitt i tabell 8.6 på neste side.

Resultatet av modellen med regresjon $\ln(\text{vekt}^2)$ på $\ln(\text{lengde}^2)$ er gitt i tabell 8.8 på side 131 og i tabell 8.9 på side 132, samt vist i figur 8.4 på side 132. I denne figuren er modellen transformert tilbake til opprinnlig skala. Resultatene med denne modellen er ikke helt sammenlignbare med modellen fastlåst i tredje potens. Dette er fordi tallene er transformert, men resultatene fra denne modellen viser også større varians hos LD24:0-gruppen (tabell 8.9 på side 132).

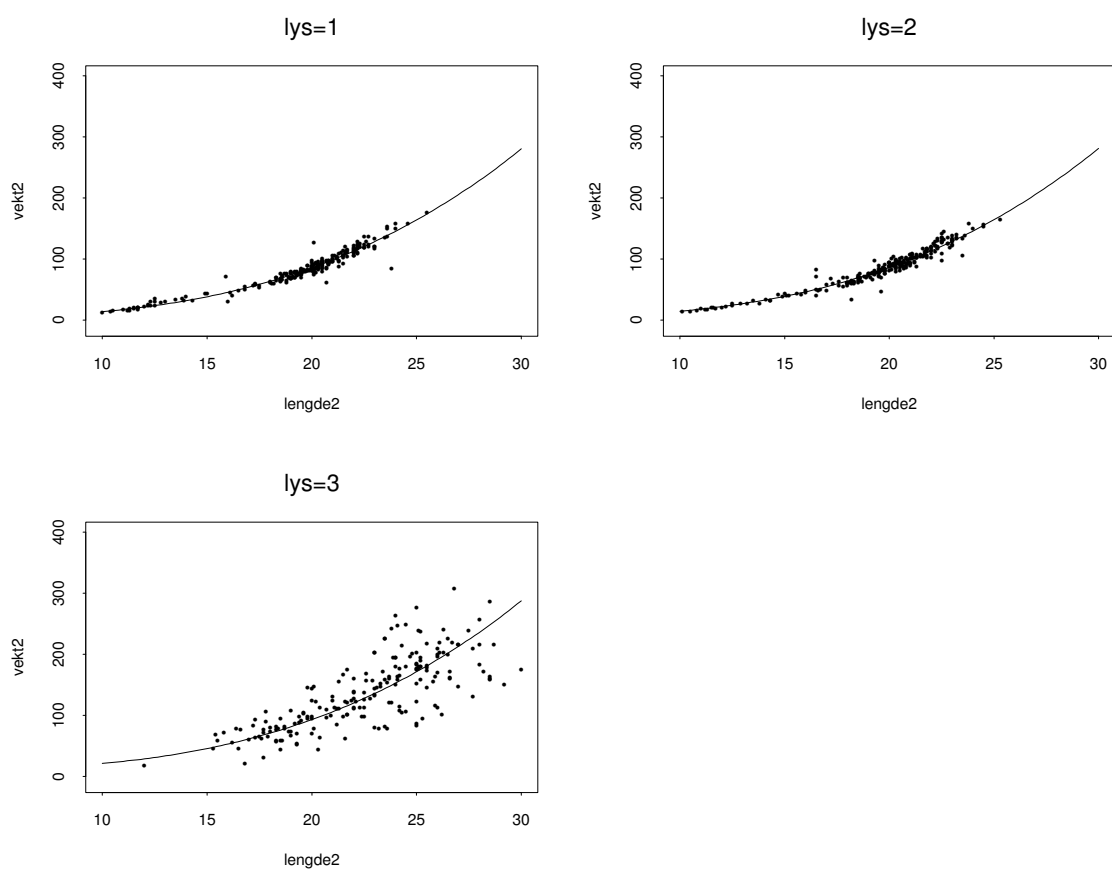
Hensikten med dette avsnittet var å påvise ulik varians i lysgruppene. Dette vil også virke inn på kovariansmatrisene ved MANOVA, MACOVA og LME. Jeg antar imidlertid at metodene er rimelig robuste mot dette. Variansen er også mest fremtredende når vi ser på forholdet mellom lengde og vekt.

Tabell 8.6: Konfidensintervall (95 prosent) for effekter, varians og residual. Modell: vekt2 forklart ved $I(len2^3)$.

interK1	Nedre	Estimat	Øvre
Effekter			
(Intercept)	1.52	3.50	5.47
lys2	-0.66	0.68	2.03
lys3	1.69	7.34	12.99
$I(len2^3)$	0.0100	0.0102	0.0105
Varians funksjon			
lys2	0.88	1.00	1.14
lys3	4.61	5.27	6.02
St. feil residual			
St. feil	6.8	7.5	8.2

Tabell 8.7: Koeffisienter: logvekt forklart ved loglengde + konstantledd.

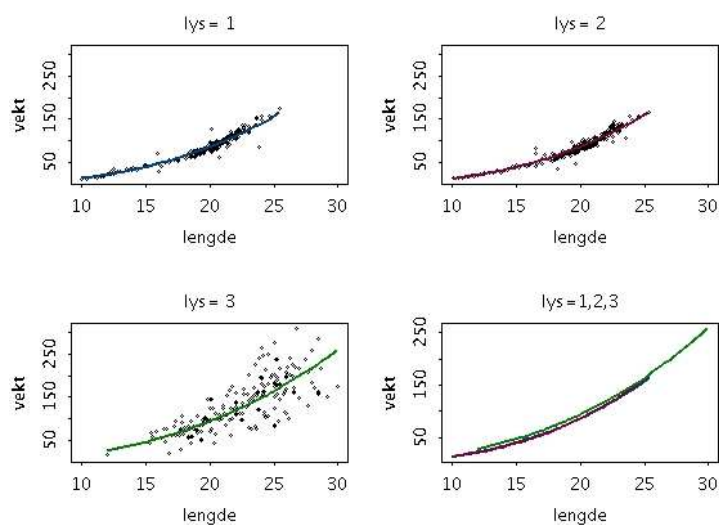
kondis2	verdi	st.feil	t-verdi	p-verdi
(Intercept)	-3.62	0.099	-36.56	< .0001
lys2	-0.30	0.15	-2.04	0.0419
lys3	0.71	0.41	1.73	0.0832
log(len2)	2.69	0.034	80.26	< .0001
lys2log(len2)	0.103	0.05	2.05	0.0403
lys3log(len2)	-0.21	0.13	-1.56	0.1191



Figur 8.3: Spredningsplott av lengde2 mot vekt2. Linjene representerer modellen. Plottet er delt opp i tre, etter lysbehandling. 1 = SNL, 2 = SNL+B og 3 = LD24:0.

Tabell 8.8: Koeffisienter: logvekt forklart ved loglengde + konstantledd.

kondis2	verdi	st.feil	t-verdi	p-verdi
(Intercept)	-3,49	0,14	-24,7	< .0001
lys1	-0,150838	0,074	-2,0	0,042
lys2	0,286554	0,13	2,1	0,033
log(len2)2,660578	0,046	58,3	< .0001
lys1log(len2)	0,051348	0,025	2,1	0,040
lys2log(len2)	-0,085732	0,043	-2,0	0,048



Figur 8.4: Spredningsplott av loglengde2 mot logvekt2. Linjene representerer modellen. Plottet er delt opp i tre, etter lysbehandling. 1=SNL, 2=SNL+B og 3= LD24:0.

Tabell 8.9: Konfidensintervall (95 prosent) for effekter, varians og residual. Modell: logvekt2 forklart ved loglengde2.

interK1	Nedre	Estimat	Øvre
Effekter			
(Intercept)	-3,8	-3,5	-3,2
lys1	-0,30	-0,15	-0,005
lys2	0,02	0,29	0,55
log(len2)	2,57	2,66	2,75
lys1log(len2)	0,0022	0,051	0,100
lys2log(len2)	-0,17	-0,086	-0,0007
Varians funksjon			
2	0,89	1,02	1,17
3	2,57	2,94	3,37
St. feil residual			
St. feil	0,088	0,097	0,107

8.1.4 Biologisk diskusjon

Vekt og lengde

Nedbøyingen mot venstre i figur 8.1 på side 123 og figur 8.2 på side 126 er forårsaket av vekststagnasjon hos de minste individene. Dette er samme fenomen som bimodalitet slik det rapportert av Skilbrei (1991) og Saunders *et al.* (1994). Vi ser av histogrammet at LD24:0-gruppen har få fisk under 50 gram, mens de to andre gruppene faktisk inneholder en del slike fisk.

At denne utflatingen ikke forekommer hos LD24:0-gruppen kan henge sammen med at denne gruppen får ett sommersignal. Vi er da inne i den klassiske diskusjonen om lyset er direkte vekststimulerende eller om det påvirker sesongavhengig vekst ().

En mulig tolkning av fenomenet er at i de gruppene der laksen har et lysregime tilsvarende naturlig lys, vil de største laksene spise mindre fordi de alt har nådd en kritisk størrelse for smoltifisering.

Vi kan tenke at LD24:0-gruppen forventer en vinter foran seg og vil derfor vokse mest mulig foran en i naturen matfattig sesong. En vanlig tolking av bimodalitet i vekst hos lakseparr er at man tolker den nedre modalen til å være parr som ikke «kan» bli smolt påfølgende vår og denne derfor bremser veksten for heller å oppnå optimal størrelse som toårig smolt. En tilsvarende tolkning vil være mulig med hensyn på at de største parrene alt har oppnådd en optimal smoltstørrelse. Siden det ikke er gjort studier av adferden til fisken i dette forsøket, blir tolkingen mer gjetning. Målet mitt er heller ikke først og fremst å si at slik er sammenhengene, men heller peke på mulige sammenhenger som man kan følge opp med eventuelle nye forsøk.

En annen sak er at samme datasett er grundig analysert av Stefansson *et al.* (1991), og jeg bare bidrar med tilleggsopplysninger i forhold til deres arbeid. De har også flere målinger i denne fasen, slik at det har liten hensikt å gå dypere inn i dette. Følgende forhold fra deres kan være viktig å ha med seg om man vil tolke ferskvannsfasen mer fullstendig:

Stefansson *et al.* (1991) har økt vanntemperaturen i forsøket sent i desember 1987 (ferskvannsfase), noe som ser ut til å ha hatt en kraftig effekt på kondisjonsfaktor til gruppen med simulert naturlig lys (SNL). Denne gruppens kondisjonsfaktor øker fra ca. 1,2 til ca. 1,35 fram til 28. januar 1988. Kondisjonsfaktoren faller deretter til i underkant av 1,1 fram til 26. april 1988 som er relativt likt de andre gruppene.

Morfologiske forandringer knyttet til smoltifisering oppsto tidligere i LD24:0 gruppen, men all fisk i forsøket er klassifisert som smolt sent i april basert på ytre kjennetegn.

«Kondisjonsfaktor»

LD24:0 ser ut til å ha en mye større varians (4–6 ganger mer) i forholdet mellom lengde og vekt (vist i figur 8.3 på side 131 og i tabell 8.6 på side 130).

Dette kan ha noe med en forvirringseffekt å gjøre, eller at konkurransen hos disse fiskene er hardere. Dette drøftes nærmere i kapitlet om kjønnsforskjeller, fordi det ser ut til at effekten er mest fremtredende hos hunnene. Jeg har ikke sett samme fenomen rapportert fra forsøk med bruk av samme lysregime. Dette kan ha sammenheng med at fordelingen i dataene sjelden vises. Like sjelden diskuteres spredningen i dataene.

Det hadde også vært interessant å se om dette er tilfellet i begge parallellene (de to karene som får lysregime LD24:0). Desverre har jeg ikke opplysninger som kan skille de to parallellene. I små kar med få fisk kan det oppstå kareffekter, der fiskene i noen kar har en helt annen «sosial oppførsel» (adferd) enn i de andre karene (personlig observasjon på Austevoll Havbruksstasjon). På den annen siden ser fenomenet ut til være knyttet til hunnene i så sterk grad at begge parallellene sannsynligvis har stor varians.

8.1.5 Vurdering av statistiske metoder

De tilpasninger jeg har gjort i ANCOVA-modellene blir ikke så trivielle som ved første øyekast i lys av Crawley (2002), Sokal og Rohlf (1981) og Snedecor og Cochran (1989). De mener at man bør bruke likt stigningstall i alle grupper, og trekker ikke inn variansfunksjoner (med unntak av Crawley (2002)). Jeg antar begrunnelsen deres for likt stigningstall er at det letter tolkingen av effekten fra kategoriske variabler. Videre kommentarer til modellene jeg har brukt er:

1. Separate modeller

- Det kan kritiseres at jeg ikke ser på vekt og lengde sammen, men for å kunne bruke en metode basert på binormalitet må vekt og lengde log-transformeres. Dessuten får jeg resultater som er mer sammenlignbare med det som ellers er gjort på vekst hos laks. En måte å kompensere for at vi da får to (tre) tester er å justere signifikansnivået tilsvarende. Dette bør gjøres når testene gjelder

samme hypotese, og ikke dersom ulike tester gjelder ulike hypoteser. Målet er ikke testing, men estimering av biologiske parametre, inkludert usikkerhet.

2. Valg av polynom

- Det er viktig å være oppmerksom på at oppbøyningen vi ser mot høyre i plottet ikke har noe med realiteten i gjøre, men bare er en konsekvens av at jeg har brukt et tredjegradspolynom som modell. Nedbøyningen mot venstre i figur 8.1 på side 123 og figur 8.2 på side 126 antas forårsaket av vekststagnasjon hos de minste individene. Dette innebærer at vi har to grupper som er med i samme regresjon.

3. Valg av variansfunksjon

- Residualplottene i figur A.1 antyder muligens at jeg burde brukt en variansfunksjon som lar variansen øke med kovariaten eller med forventet verdi. Men mønsteret er ikke like entydig innen alle gruppene.

4. Valg av modell for «kondisjonsfaktor»

- Jeg angriper forholdet mellom lengde og vekt på en litt annen måte enn tradisjonelt innen fiskebiologi, der man ofte anvender Fultons kondisjonsfaktor. Ved å se på forholdet mellom lengde og vekt i spredningsplott blir de ulike variansene meget tydelig. De to modellene er ikke innbyrdes konsistente med hensyn på feilledet. Ved begge modeller får jeg estimert ulike varianser hos lysgruppene.

5. Residualplott

- Normalplottene (for vekt) i figur A.1 antyder noe avvik fra normalfordeling i halene, men ikke nok avvik til at vi bryr oss med det. Vi ser en opphoping av datapunkter rundt 100 gram i residualplottene for SNL- og SNL+B-gruppen. Dette kommer fordi modellen flater ut, slik at det ikke predikeres større verdier.
- Normalplottene (for lengde) i figur A.2 antyder også noe avvik fra normalfordeling, spesielt for nedre hale i lysregime SNL+B. Vi ser også her en opphoping av datapunkter i residualplottene for SNL- og SNL+B-gruppen.
- Normalplottene (for vekt² forklart med lengde²) i figur A.3 antyder avvik fra normalfordeling i ekstrem grad hos LD24:0-fisken. Blant annet derfor prøvde jeg en ekstra modell. Men hovedfokuset mitt er på de ulike variansene.

«Støyen» (dvs. residualene fra modellen) kommer i hovedsak fra individuelle forskjeller mellom fiskene, og etter min oppfatning «tilfeldigheter» i selve fôrings situasjonen, der plassering i karet og hiarki spiller inn. At modellen ikke er perfekt kan gi bidrag til residualene.

9 Kjønn

Problemstilling

Jeg ser her på om det er forskjeller i vekt hos hanner og hunner, uten å ta hensyn til eventuell modning. Lys tar jeg dermot hensyn til som faktor. Ferskvannsfasen og sjøfasen kan være biologisk interessante hver for seg. Ved å gjøre denne oppsplittingen får jeg og en utdyping av resultatene fra ANCOVA-modellen i forrige kapittel, der kjønn ikke var med.

Metoder

Jeg bruker ulike statistiske metoder i de ulike fasene. I ferskvannsfasen bruker jeg en ANCOVA-modell, mens i sjøfasen bruker jeg MANOVA og MANCOVA. Det er slik at andre måling i mindre grad enn senere målinger oppfyller kravene til fordeling både for MANOVA og MANCOVA. Derfor kan det være gunstig å bruke en ANCOVA-modell i ferskvannsfasen.

Jeg bruker en ANCOVA-modell tilsvarende den jeg har brukt for vekt i forrige kapittel, men jeg legger inn mulighet for vertikal forskyvning av kurver avhengig av kjønn.

Når kjønn er med som faktor blir det færre fisk som kan være med i analysen. Dette påvirker resultatene. Derfor har også kjørt en ANCOVA-modell på samme datasett som jeg bruker når kjønn er med. Dette er gjort får å vise i hvilken grad resultatene blir påvirket av færre målinger, og de blir faktisk ikke det i nevneverdig grad. Resultatene fra denne analysen er kun vist i appendiks B.

Jeg bruker en ANCOVA-modell tilsvarende den førstnevnte men med rom for ulike varianter for alle kombinasjoner av lys og kjønn.

Innsnevring

Når jeg sammenligner de statistiske metodene er det best at dette gjøres på en komplett del av datasettet. Med det mener jeg at det ikke kan være manglende målinger på enkelteindivid på noe tidspunkt (måling 3–6).

For å få en mer ensrettet oppgave ser jeg ikke på de to siste målingene hverken angår lys, kjønn eller modning. Å ikke se på de to siste målingene reduserer også oppgavens omfang. Jeg ser heller ikke på lengde, også for å begrense oppgavens omfang.

9.1 Ferskvannsfasen

9.1.1 Vekt

Jeg har brukt samme modell som i ferskvannsfasen i forrige kapittel (tredjegradspolynom og varIdent), men med kjønn som ekstra faktor (uten samspill). Jeg har fått følgende resultater gitt ved tabell 9.1 og Splus-utskrift i appendiks B.

Tabell 9.1: Koeffisienter: ANCOVA-modell: tredjegrads-polynom, vekt2 forklart med lys og kjønn.

sexV2	verdi	st.feil	t-verdi	p-verdi
(Intercept)	94.4	2.50	37.8	< .0001
kjønn	0.15	2.22	0.068	0.9455
lys2	0.82	2.48	0.33	0.7412
lys3	53.9	3.29	16.36	< .0001
vektsentrert1	2.8	0.47	6.08	< .0001
$I(\text{vektsentrert1}^2)$	-0.47	0.07	-6.8	< .0001
$I(\text{vektsentrert1}^3)$	0.0236	0.006	4.1	0.0001
lys2vektsentrert1	0.98	0.51	1.9	0.0578
lys3vektsentrert1	6.2	0.73	8.5	< .0001

Tabell 9.1 viser at kjønn ikke justerer modellen vertikalt noe av betydning (0,15 gram). P-verdien på 0,94 sier at dette kan være rent tilfeldig. Resten av parametrene er konsistente med modellen som ikke tar hensyn til kjønn.

Men det er noen forandringer i estimatene, for eksempel for stigningstallet innen lys₃ (LD24:0) i forhold til tilsvarende modell i forrige kapittel. Dette kommer av at datasettet er redusert når vi tar hensyn til faktoren kjønn. Dette ser vi av tabell B.1 på side 179 og Splus-utskrift i appendiks B. Jeg bruker da samme de dataene som i modellen tidligere men uten kjønn som faktor. Dette viser at den innflytelse kjønn ser ut til å ha på tilpasningen av modellen antagelig kommer av at datasettet er redusert.

Ved kjøring av modellen med kjønn men med ulike varianser for alle kombinasjoner av lys og kjønn, får jeg at kjønn ikke påvirker variansen. Dette er vist i Splus utskrift i appendiks B, og resultater i tabell B.2 på side 180, også i appendiks B.

Jeg velger å ikke gå videre med mer kompliserte modeller, der jeg for eksempel har samspill mellom kjønn og andregradsleddet.

9.2 Sjøvannsfasen

Jeg ser her på forskjeller hos kjønnene, der vi tar hensyn til lys, men ikke tol modning. Jeg bruker først MANOVA, så MANCOVA. Dette for å få sammenlignet metodene.

9.2.1 Vekt

MANOVA

Vi starter med samspill mellom lys og kjønn. Splus-utskriften i appendiks B og tabell B.3 på side 181, også i appendiks B viser at samspillet mellom lys og kjønn ikke er statistisk signifikant. (I appendiks B viser jeg også «korrelasjonsstrukturen» ved figur B.1 på side 182, tabell B.4 på side 182 og tabell B.5 på side 183)

Jeg prøvde en MANOVA-modell uten samspill, og fikk da heller ikke vesentlige kjønnsforskjeller. Dette er vist i tabell 9.2 på neste side og tabell 9.3 på neste side.

MANCOVA

Jeg prøver en MANCOVA med startvekt som kovariat og antar en additiv modell med vekt1 (startvekt), lys og kjønn som forklaringsvariabler. Tabell 9.4 på neste side og tabell 9.5 på side 141 viser igjen at kjønn ikke har effekt i modellen.

Tabell 9.2: Testobservatorer og tilhørende p-verdier for vekt3-6 (f-verdi er en tilnærming). MANOVA-modell: lys + kjønn (data= smolt.manova.alle4, contrasts = list(kjønn = contr.treatment)).

manNKMV1tab	estimat	f-verdi	frihetgr. teller	frihetgr. nevner	p-verdi
Pillai Trace					
lys	0,156	7,7	8	728	0
kjønn	0,01	0,879	4	363	0,477

Tabell 9.3: Koeffisienter: vekt 4–6: MANOVA-modell:
Forklaringsvariabler: lys+kjønn.

ancovaVuS	vekt3	vekt4	vekt5	vekt6
(Intercept)	403,95	987,28	1219,69	1636,27
lys1	-10,99	-4,44	4,34	8,93
lys2	18,28	8,37	-2,28	-25,08
kjønn	3,81	13,00	10,93	-27,57

Tabell 9.4: Testobservatorer og tilhørende p-verdier for vekt3-6 (f-verdi er en tilnærming). MANCOVA-modell: vekt1 + lys + kjønn. (data= smolt.manova.alle4, contrasts = list(kjønn = contr.treatment)).

tab:mancLKadst	estimat	f-verdi	frihetgr. teller	frihetgr. nevner	p-verdi
Pillai Trace					
vekt1	0,226	26,5	4	362	0
lys	0,18	9,0	8	726	0
kjønn	0,011	1,0	4	362	0,41

Tabell 9.5: Koeffisienter: vekt 4–6: MANCOVA-modell.: Forklaringsvariabler: vekt1+lys+kjønn.

ancovaVuS	vekt3	vekt4	vekt5	vekt6
(Intercept)	266,39	703,443	876,5	1197,89
vekt1	12,00	24,765	29,9	38,25
lys1	−8,42	0,866	10,8	17,14
lys2	24,52	21,264	13,3	−5,16
kjønn	8,88	23,466	23,6	−11,41

9.3 Diskusjon

9.3.1 Biologisk diskusjon

Jeg fant bare marginale vektforskjeller mellom kjønnene. Dette er i overensstemmelse med funn hos Nævdal *et al.* (1978b).

Sett i forhold til at hanner og hunner forventes å ha litt ulike strategier med hensyn på modning, hadde jeg ventet å finne større forskjeller mellom kjønnene. Jeg har i denne oppgaven imidlertid ikke sett på de to siste målingene, og det er mulig at veksten der er mer ulik. Man kan også si at vekstforskjellene kommer til uttrykk når man ser på moden kontra umoden fisk. Se kapittel om modning (kap. 10) for videre diskusjon.

Men jeg finner betydelige forskjeller mellom kjønnene i forholdet mellom lengde og vekt innen lysregime LD24:0. Dette diskuteres mer i neste avsnitt.

Samspill: lys og kjønn?

Jeg har funnet betydelige forskjeller bare mellom kjønnene innen lysregime LD24:0, slik det er vist i kapittelet 6.4. Forskjellene fremtrer når man ser på «forholdet» mellom lengde og vekt. Jeg har ikke gjort testing på at variansene er ulike, men mener det trer tydelig fram på figur 6.12 på side 70 og de etterfølgende figurer.

Oppedal *et al.* (2001) refererer å finne endret atferd ved lysbehandlig i merder. De ser at på dagtid svømmer fisken rundt i «stim», mens dette mønsteret oppløser seg om natten. Ved å bruke ekstra lys om natten observeres at fisken svømmer rundt i «stim» også da. Det

er mulig at LD24:0-fisken har hatt tilsvarende atferd, men det forklarer ikke hvorfor bare hunnene har fått svært varierende «kondisjonsfaktor». Den kontinuerlige svømmingen kunne tenkes å forsterke delingen i vinnere og tapere, altså høy og lav kondisjonsfaktor.

Det kan altså være konkurranse om føret som avgjør kondisjonsfaktoren, eller at fisken (hunnene) i denne gruppen er forvirret på grunn av den lange perioden med kontinuerlig lys. Vi er kanskje utenfor fiskens normale område for reaksjonsnormer, og de enkelte fisk reagerer svært ulikt. Grunnen til at hunnene i LD24:0-gruppen reagerer slik er ikke opplagt.

Jeg har også vurdert om dette kunne komme fra «feil i filene», men mener at dette er urealistisk (lengde og vekt til samme individ henger sammen under filbehandling).

9.3.2 Vurdering av statistiske metoder

Se kapittel om modning, der de samme metodene diskuteres.

10 Modning

Problemstilling

I dette kapitlet ser jeg på forskjeller i lengde og vekt mellom tert og laks som modner senere. Til dette formål bruker jeg MANOVA og MANCOVA. Jeg tar ikke hensyn til lys som faktor. Grunnen til det at jeg ellers får for lite antall i hver «celle».

En annen hypotese enn om lys er med, blir da testet. Man undersøker da om tertene har en annen størrelse enn senere moden laks. Vi ser altså bort fra lysregimes påvirkning på størrelse. Denne hypotesen er like relevant som om lys var med.

Metoder

Jeg bruker modning nøstet innen kjønn. Jeg siterer Pearce(1992) i Nester (1996) «In biological context interactions are common, so it is better to play safe and regard any appreciable interaction as real whether it is significant or not.»

Velger man å tro at det til en viss grad er tilstede et samspill kan man likegodt bruke nøsting. Generelt bruker jeg en modell med modning nøstet innen kjønn, men jeg viser også bruk av én modell med samspill. LME blir behandlet i et eget appendix, fordi metoden ikke tilfører noe ekstra rent biologisk, og for å gjøre oppgaven mer oversiktlig.

Innsnevring

Jeg har snevret inn undersøkelsen til å gjelde måling nunnene 3 til 6. dvs. perioden september 1988 og mai 1989. Antall tidspunkt jeg undersøker blir da fire.

Å kutte ut de to først kan være diskutabelt. Men som nevnt tidligere har jeg bare noen av målingene for ferskvannsfasen, og ser ingen grunn til gå grundig til verks der jeg har ett mangelfullt materiale.

Tanken med å kutte ut de to første målingene var også å snevre inn antall variabler for at effektene ikke skal drukne i ikke-effekter. Muligens taes «beslutningen» om å modne som tert i sjøvannsfasen. Da er det veksten i denne fasen som er relevant. Dette blir diskutert i slutten av dette kapitlet. Ved MANCOVA-metoden blir de to første variablene tatt med som forklaringsvariabler.

Jeg kutter også trinnvis ut et eller flere tidspunkt fra modellen, som en eliminasjonsmetode for å få lokalisert hvilket måletidspunkt som gjør utslag i testen. For å få sammenlignbare resultater når jeg kutter ut et og et måletidspunkt må jeg ha komplette måleserier. Jeg ser derfor kun på de individene som ikke inneholder manglende verdier for vekt3-6 og len3-6. Dette gjør datasettet mer balansert. Jeg antar også at Splus må ha komplette måleserier for å fungere «korrekt» i MANOVA-rutinen.

Jeg har også kuttet ut de postsmoltmodne, som er så få fisk at de ikke passer inn i en MANOVA/MANCOVA .

10.1 MANOVA

Hver celle inneholder flere observasjoner enn antall variabler, og jeg bruker totalt 370 fisk. Jeg ser først på vekt, nærmere bestemt vekt3-6, deretter på lengde3-6.

	Hanner	Hunner	
Umoden	138	198	336
Moden	25	9	34
	163	207	370

Tabell 10.1: Fordeling av modne og umodne hanner og hunner.

Tabell 10.2: Testobservatorer og tilhørende p-verdier for vekt3-6 (f-verdi er en tilnærming). MANOVA-modell: kjønn*modn.

manNKMVtab	Estimat	f-verdi	frihetgr. teller	frihetgr. nevner	p-verdi
Pillais spor					
kjønn	0,0091	0,8323	4	363	0,5052
modn	0,0912	9,107	4	363	0
kjønn:modn	0,0162	1,4936	4	363	0,2035

Tabell 10.3: Effekter på vekt. MANOVA-modell: kjønn i samspill med modning.

ManovaNV	vekt3	vekt4	vekt5	vekt6
(Intercept)	403,5	993,2	1223,8	1603,9
kjønn	-1,3	6,9	8,28	8,8
modn	-21,3	-31,9	7,93	198,1
kjønn:modn	6,3	-28,9	-23,7	-97,7

10.1.1 Vekt

Samspill

Tabell 10.2 viser ikke signifikans på samspill, og tabell 10.3 viser estimerte effekter av kjønn og modning. Begge tabellene indikerer at kjønn er av marginal betydning og at modning har en effekt.

Nøsting

Tror jeg på samspill, kan jeg like godt kjøre nøsting. Jeg bruker så en modell med modning nøstet innenfor kjønn. Resultatene er vist i tabell 10.4 på neste side og tabell 10.5 på neste side. Resultatene er like som om vi bruker samspill. Dersom man «legger til samspilleffekten» får man de samme tallene som ved nøsting.

Alle testobservatorene i tabell 10.4 gir omtrent samme resultat. Dette tyder på uvesentlige kjønnsforskjeller, men derimot en effekt av modning innen kjønn. Eventuelle forskjeller

Tabell 10.4: Testobservatorer og tilhørende p-verdier for vekt3-6 (f-verdi er en tilnærming). MANOVA-modell: kjønn/modn.

manNKMVtab	Estimat	f-verdi	frihetgr. teller	frihetgr. nevner	p-verdi
Pillais spor					
kjønn	0,009	0,832	4	363	0,505
modn innen kjønn	0,107	5,134	8	728	0
Wilks lambda					
kjønn	0,991	0,832	4	363	0,505
modn innen kjønn	0,894	5,21	8	726	0
Roys største rot					
kjønn	0,009	0,832	4	363	0,505
modn innen kjønn	0,104	9,493	4	364	0

Tabell 10.5: Effekter i MANOVA-modell med modning nøstet innen kjønn.

ManovaNV	vekt3	vekt4	vekt5	vekt6
(Intercept)	404	993	1220	1600
kjønn	-1,25	6,94	8,28	8,84
kjønn1modn	-2,76	-3,08	31,6	296
kjønn2modn	-14,9	-60,8	-15,7	100

er mindre enn en 0,001 i p-verdi. F-observatoren for Roys største rot er større enn for Pillais spor og Wilks lambda. Dette kan tyde på at effekten er lokalisert på få tidspunkt.

Det ser også ut til i tabell 10.5 at vi har størst effekt på modning for hannene. Videre kan det se ut til at effekten er lokalisert til siste måling ut fra effektstørrelsen.

Elimineringsmetode

For å lokalisere hvor effekten er har jeg brukt en slags elimineringsmetode. Resultatene fra elimineringsmetoden er vist i tabell 10.6 på neste side.

Tabell 10.6: Testobservatorer og tilhørende p-verdier for «elimineringemetoden» vekt3-6. MANOVA-modell: kjønn/modn.

manNKMVtab	Estimat	f-verdi	frihetgr. teller	frihetgr. nevner	p-verdi
cbind(vekt3, vekt4, vekt5)					
kjønn	0,0023	0,27	3	364	0,84
modn in kjønn	0,035	2,2	6	730	0,039
cbind(vekt3, vekt4)					
kjønn	0,0015	0,27	2	365	0,76
modn in kjønn	0,013	1,15	4	732	0,33
cbind(vekt4, vekt5)					
kjønn	0,001	0,18	2	365	0,83
modn in kjønn	0,027	2,5	4	732	0,042
cbind(vekt5, vekt6)					
kjønn	0,0072	1,3	2	365	0,27
modn in kjønn	0,0796	7,6	4	732	0

For å lokalisere det eller de tidspunkt som slår ut, eliminerte jeg ett og ett tidspunkt av gangen. Jeg henviser til tabell 10.6 på neste side. Jeg kuttet først ut vekt6 og så vekt5.

Ved å kutte ut vekt6 fikk jeg fremdeles lav p-verdi på modning innen kjønn, men ved å kutte ut vekt5 så fikk jeg ingen «sikre effekter». Jeg kuttet så ut vekt3 og beholdt kun vekt4 og vekt5. Dette ga bare en marginal forandring av p-verdien for modning innen kjønn, altså gir vekt3 intet bidrag til den lave p-verdien. For kompletthetens skyld og for å ta vekk effekten av vekt4, gjorde jeg en kjøring uten vekt4, dvs. bare med vekt5 og vekt6:

I forhold til modellen med alle tidspunkt så har F-observatoren økt for modning innen kjønn (p-verdien får vi dessverre ikke se da den er mindre enn 1/10000). Jeg konkluderer med at effekten er på vekt5 og vekt6 og at de modne har vokst litt mer i perioden mellom måling 4 og 5, men at effekten er størst mellom måling 5 og 6 (se diskusjon for utdyping).

Forutsetninger

Forutsetningen om normalfordeling og multinormalfordeling ser ut til å være rimelig bra oppfylt. Målingene viser seg å være tilnærmet normalfordelt ut i fra normalplottene av residualene vist i figur C.1.1 på side 186. Spredningsplottene vist i figur C.1.1 på side 184 gir inntrykk av linearitet, som kan støtte opp under antagelsen om binormalitet. I det samme appendikset har jeg også vist korrelasjonsmatrisen og kovariansmatrisen i tabell C.1 på side 185 og i tabell C.2 på side 185. Korrelasjonsmatrisen viser at korrelasjonen dempes med tiden, men at den er svært høy i vintersesongen da fisken vokser lite. Kovariansmatrisen viser også at variansen øker med tiden.

10.1.2 Lengde

Jeg har her sett på lengde, nærmere bestemt len₃₋₆. Jeg «kjørte» også her nøsting, altså en modell med modning nøstet innenfor kjønn. Resultatene vist i tabell 10.7 på neste side og tabell 10.8 på neste side ble relativt like resultatene for vekt med samme metode. Det ser ut til at vi har størst effekt på modning for hannene.

Jeg har også vist korrelasjonsmatrisen og kovariansmatrisen i tabell C.3 på side 185 og i tabell C.4 på side 187. Korrelasjonsmatrisen viser at korrelasjonen dempes med tiden også for lengde, men at den er svært høy i vintersesongen da fisken vokser lite. Kovariansmatrisen viser at variansen øker med tiden, men i en helt annen størrelsesorden enn for lengde. Variansen dobles nesten for lengde, mens for vekt økte den med en faktor på 10.

For å lokalisere det eller de tidspunkt som slår ut, eliminerte jeg et og et tidspunkt av gangen. Jeg kuttet først ut len₆ og så len₅. Altså samme prosedyre som for vekt. Resultatene er vist i tabell 10.9 på neste side.

Ut fra tabell 10.9 på forrige side ser effekten ut til å ligge hos len₅ og len₆, derfor kjørte jeg en modell med bare len₅₋₆. Jeg konkluderer med at mønsteret er forholdsvis likt for lengde som for vekt. Jeg har også sett på vekt og lengde sammen i en modell. Men siden hensikten med denne oppgaven er å etablere metodene og å se at de virker, heller enn å gjennomanalysere materialet velger jeg ikke å gå nærmere inn på dette.

Tabell 10.7: Testobservatorer og tilhørende p-verdier for lengdet3-6 (f-verdi er en tilnærming). MANOVA-modell: kjønn/modn.

manNKMVtab	Estimat	f-verdi	frihetgr. teller	frihetgr. nevner	p-verdi
Pillais spor					
kjønn	0,0156	1,44	4	363	0,22
modn in kjønn	0,046	2,14	8	728	0,030
Roys største rot					
kjønn	0,0159	1,44	4	363	0,22
modn in kjønn	0,0345	3,14	4	364	0,015

Tabell 10.8: Effekter på lengde3-6 i MANOVA-modell med modning nøstet innen kjønn.

ManovaNL	len3	len4	len5	len6
(Intercept)	32,82	42,618	46,29	50,699
kjønn	0,12	0,162	0,12	0,028
kjønn1modn	-0,18	0,503	0,97	1,910
kjønn2modn	-0,67	-0,058	-0,52	0,439

Tabell 10.9: Testobservatorer og tilhørende p-verdier for «elimineringmetoden» lengde3-6. MANOVA-modell: kjønn/modn.

maneliNL	Estimat	f-verdi	frihetgr. teller	frihetgr. nevner	p-verdi
cbind(len3, len4, len5)					
kjønn	0,0107	1,31	3	364	0,271
modn in kjønn	0,0314	1,94	6	730	0,0722
cbind(len3, len4)					
kjønn	0,0015	0,28	2	365	0,7559
modn in kjønn	0,0153	1,41	4	732	0,2279
cbind(len5, len6)					
kjønn	0,0046	0,85	2	365	0,4291
modn in kjønn	0,0219	2,02	4	732	0,0893

Tabell 10.10: Testobservatorer og tilhørende p-verdier for vekt3-6 (f-verdi er en tilnærming). MANCOVA-modell: vekt1 + kjønn/modn. (data=smolt.manova.alle4, contrasts = list(modn = contr.treatment)).

manNKMV1tab	Estimat	f-verdi	frihetgr. teller	frihetgr. nevner	p-verdi
Pillais spor					
vekt1	0,2204	25,59	4	362	0
kjønn	0,0095	0,87	4	362	0,48
modn in kjønn	0,11	5,2687	8	726	0

10.2 MANCOVA

Som nevnt i kapittelet 7. gir MANCOVA svar på andre hypoteser enn MANOVA. Spørsmålstillingen blir da om de modne vokser bedre kompensert for startvekt, eventuelt om de over hele forsøksperioden har vokst mer (eller mindre dersom det er hypotesen). For å finne ut dette kjørte jeg først en MANCOVA med startvekt som kovariater, deretter smoltvekt, etc. (sentrering av kovariat var ikke nødvendig da Splus ordnet dette selv). Modning var som ellers nøstet innen kjønn. Antall frihetgrader minker med en til 365 da vi har en forklaringsvariabel mer enn ved MANOVA.

10.2.1 Vekt

Ut i fra tabell 10.10 og tabell 10.11 på neste side ser dette ut til å gi lignende resultater MANOVA-modellen for vekt. Veksten i fra måling 5 til 6 til de modne ser ut til være bedre enn hos umodne fisk, og en tendens det samme fra måling 4 til 5.

Modellen med smoltvekt som kovariat ser ut til å gi litt andre resultater ut fra tabell 10.13 på neste side. Veksten i fra måling 3 til 4 hos hannene til de modne ser ut til være bedre enn for umodne hanner, og en antydning til det samme fra måling 4 til 5 hos hunnene.

10.2.2 Lengde

Her ser vi på lengde3-6 ved hjelp av mancova, der lengde1 er kovariat. Jeg bruker Pillais spor og Roys største rot som testobservatorer. Begge observatorene antyder effekt av

Tabell 10.11: Effekter i MANCOVA-modell med modning nøstet innen kjønn. Startvekt som kovariat.

MacovaNVst	vekt3	vekt4	vekt5	vekt6
(Intercept)	276,9	720,5	891,8	1180,1
vekt1	11,3	24,3	29,5	37,7
kjønn	0,2	10,1	12,2	13,8
kjønn1modn	-37,6	-24,8	5,1	262,0
kjønn2modn	-40,7	-116,3	-83,3	14,1

Tabell 10.12: Testobservatorer og tilhørende p-verdier for vekt3-6 (f-verdi er en tilnærming). MANCOVA-modell: vekt2 + kjønn/modn. (data= smolt.manova.alle4, contrasts = list(modn = contr.treatment)).

manNKMV2tab	Estimat	f-verdi	frihetgr. teller	frihetgr. nevner	p-verdi
Pillais spor					
vekt2	0,658	161,059	4	335	0
kjønn	0,011	0,899	4	335	0,465
modn innen kjønn	0,1	4,401	8	672	0

Tabell 10.13: Effekter i MANCOVA-modell med modning nøstet innen kjønn. Smoltvekt som kovariat.

MacovaNVsm	vekt3	vekt4	vekt5	vekt6
(Intercept)	147,43	521,82	673,26	986,52
vekt2	2,46	4,50	5,23	5,84
kjønn	2,19	14,58	17,81	21,54
kjønn1modn	13,73	83,24	135,92	420,23
kjønn2modn	9,07	3,98	47,21	149,49

Tabell 10.14: Testobservatorer og tilhørende p-verdier for lengde3-6 (f-verdi er en tilnærming).MANCOVA-modell: len1 + kjønn/modn. (data=smolt.manova.alle4, contrasts = list(modn = contr.treatment)).

manNKMV1tab	Estimat	f-verdi	frihetgr. teller	frihetgr. nevner	p-verdi
Pillai Trace					
len1	0,27	33,78	4	362	0
kjønn	0,016	1,50	4	362	0,2
modn in kjønn	0,05	2,28	8	726	0,021
Roy Largest					
len1	0,37	33,78	4	362	0
kjønn	0,017	1,50	4	362	0,2
modn in kjønn	0,035	3,17	4	363	0,014

Tabell 10.15: Effekter i MANCOVA-modell med modning nøstet innen kjønn. Lengde1 som kovariat.

MacovaNV11	len3	len4	len5	len6
(Intercept)	21,36	29,75	33,47	37,52
len1	0,12	0,13	0,13	0,14
kjønn	0,13	0,17	0,12	0,03
kjonn1modn	-0,42	0,23	0,70	1,63
kjonn2modn	-1,66	-1,16	-1,62	-0,69

modning innen kjønn (tabell 10.14). Effektene er vist i tabell 10.15. Jeg viser heller ikke her diagnostiske plott.

10.3 Statistikk diskusjon

På grunnlag av bruken av metodene fører jeg videre diskusjonen fra kapittel 7. Metodene er delvis diskutert i kapittel 7, slik at dette bare er en utfyllende diskusjon.

Jeg var interessert i om metodene ga konsistente resultater, og om enkelte av metodene ga «bedre» eller avvikende resultater. Med bedre kan man mene smalere konfidensintervaller, eller at metoden fanger opp en effekt som andre metoder ikke gjør. Dette burde vært gjort på ett simulert datasett slik at vi vet hva som er «fasiten».

Oppsplitting

Det kan fremstå som «rart» at jeg først ser på kjønn nøstet innen lys i en analyse og så ser på modning nøstet innen kjønn. Intensjonen var å kjøre alle faktorene i en modell. Jeg undersøkte antallet i hver celle og så at dette ble svært lavt. Det ble for få modne av hvert kjønn oppdelt i lysgruppe. Jeg fikk betenkeligheter og delte analysen opp i to trinn, slik at jeg da hadde nok i hver celle (se avsnitt 7. for krav til antall). Dette blir da en slags nødløsning.

Virker så denne måten å gjøre det på ? Jeg tror det fordi antallet hanner, hunner og modne ikke er så ulikt innen hvert lysregime. Hadde det vært mange flere modne i gruppen med LD24:0 ville man fått en størrelseseffekt på de modne derfra.

Det ser heller ikke å være vesentlige kjønnsforskjeller hverken når man tar hensyn til lys eller ikke. Her igjen kommer samme type argument. Innenfor lysregimene er antallet av hvert kjønn noenlunde likt. Jeg påviser marginale kjønnsforskjeller innen lysregimene, da blir det også slik i et sammenslått materiale.

Hovedargumentet til støtte for se på modning nøstet innen kjønn blir at det er en relevant hypotese å undersøke uten å ta hensyn til lysregime. I et beslutningsvindu som man antar er i sjøfasen tror man at det er vekt, lengde eller energinivå som er avgjørende. Da er ikke status til fisken i ferskvann av betydning men kun status i beslutningvinduet.

Tror man på Thorpe (1994) sin hypotese om at en fisk kan ikke vite om sin størrelse, lengde eller mengde fett deponert på kroppen vil undersøkelse av veksthastighet være viktigst. Da vil lysregime være relevant fordi det antas å påvirke veksthastigheten/metabolismen også i sjøvannsfasen. Biologisk teorier er med andre ord med på å avgjøre om en statistisk metode er relevant eller ei.

Valg mellom MANOVA og MANCOVA

Dette er diskutert i kapittel 7. men jeg vil igjen peke på den biologiske problemstillingen er forskjellig for de to metodene. MANCOVA tester forskjell i variabelen kompensert for tidligere status ved bruk av en indre kovariat. MANCOVA kan og oppfattes som en ANCOVA bare at jeg ser på flere tidspunkt samtidig, og forutsetningen er strengere. Begge metoden er forøvrig relevante men de gir svar på ulike spørsmål som nevnt i kapittel 7. Dette drøftes videre i 10.4.

Forutsetninger

Forutsetninger for metoden er diskutert i resultatdelen. Jeg vil imidlertid påpeke at forutsetningene om normalfordeling holder fordi nedre modal er kuttet ut fra det materialet som studeres her.

Valg av tilnærming

Alternative tilnærminger kunne vært å bruke flere ANCOVA-modeller. Jeg ville da kunne tilpasse modellen mer enn ved MANCOVA, men dette gir mange separate tester. Alternative kontraster for MANOVA er repeterte målinger (Johnson og Wichern 2002, s. 278-282) eller profil-analyse (Johnson og Wichern 2002, s. 318-322). Dersom man velger kan man likevel sitte igjen med spørsmål om hvilket tidspunkt som gjorde profilene ulike.

Stefanssons tilnærming

Jeg vil her kort kommentere noen poeng fra Stefansson *et al.* (1991) sitt arbeid på samme datasett.

Stefansson *et al.* (1991) beregner SGR basert på gjennomsnittslengder på ulike tidspunkt. Man vil da ikke nødvendigvis få like resultater som når SGR er regnet ut for hvert enkelt individ, og at man deretter tar gjennomsnittet av de individuelle SGR.

Istedenfor å basere seg på å måle alle individene tar de et tilfeldig utvalg på ca. 100 fisk på hvert måletidspunkt. Dette kan forklare forskjellene mellom deres funn og mine.

Ved å måle så få fisk innfører de en ekstra feilkilde, kontra å måle alle fiskene. Når de da ser på utviklingen i lengde og kondisjonsfaktor, blir denne feilkilden enda mer tydelig, fordi utvalget kan være mer eller mindre representativt på de ulike tidspunkt. Dette illustrerer ulempen ved å gjøre ulike tilfeldige utvalg for hvert måletidspunkt. Størrelsen på utvalgene vil være avgjorde for om denne metoden likevel er god nok.

Størrelsen på utvalgene relativt til totalantallet vil påvirke «uavhengigheten» mellom måletidspunktene. De ser kun på de tre ulike lysregimene som forklaringsvariabler, og har således behov for færre fisk enn meg som ihvertfall har minst 12 ulike grupper. Uansett vil det gi sikrere informasjon å bruke individmerket fisk. Deres måte å gjøre dette på ble nok regnet som «state of the art» før man for alvor begynte å se fordelene med individmerket fisk.

10.4 Biologi diskusjon

Siden dette er et hovedfag i statistikk, og ikke i biologi, må denne diskusjonen ikke sees på som grundig eller komplett. Den er mer til for at statistikken ikke skal bli stående i et tomrom, og kan heller sees på som et forarbeid til en mer seriøs drøfting. Siden jeg ikke har undersøkt alle forhold vedrørende materialet, blir diskusjonen også innskrenket til noen få problemstillinger.

Mønsteret i korrelasjonmatrisene med minkende korrelasjon med tiden og en høy korrelasjon mellom målingene i vintersesongen finner man også hos Nævdal *et al.* (1978b) (i tabell 1.). De har standardisert dataene relativt til avstamming, mens jeg ser på residualer fra modeller som tar hensyn til enten lys og kjønn eller kjønn og modning.

Alle metodene antyder forskjell i vekstmønster hos tert og MSW-laks. Generelt kan vi si at tert vokser bedre enn MSW-laks i middelvei, men at spredningen er stor innen begge grupper. Dette er i overenstemmelese blant annet med Hunt *et al.* (1982) og Kadri *et al.* (1996). Hunt *et al.* (1982) finner også forhøyet vekst hos modnende tert-hanner vinter/vår og topper av nivået av 11-oxotestosterone.

Metodene antyder litt ulike resultater. MANOVA tyder på økt vekt hos terten i mai og muligens i mars. MANCOVA med startvekt som kovariat, gir omtrent samme resultater. Ser vi på MANCOVA med smoltvekt som kovariat så får vi resultater som tyder på at terten har vokst bedre i hele sjøfasen (de målingene som analyseres). Dette stemmer forsåvidt bra med Thorpe (1994) sitt forslag om at det eksisterer en terskel på veksthastighet istedenfor på størrelse/fettnivå.

Alle metodene antyder også litt forskjell mellom modne hanner og hunner, selv om de i hovedsak følger samme mønster. Hannene som modner som tert, har et tydeligere adskilt mønster. De har en sterkere vektøkning, og det ser til at denne starter tidligere enn hos hunnene. Hunnene som modner som tert har for såvidt det samme mønsteret som hannene men det er mindre tydelig og de utvikler økt størrelse senere i tid. Vi har svært få terthunner, slik at resultatene her er mer usikre.

Når det gjelder den høyere vekstraten og dermed økte størrelsen hos tertene er det imidlertid vanskelig å si om den skyldes at tertene modner eller om det er dette som får tertene til å modne (fenotypisk plastisitet).

I det første tilfellet kan vi tenke at fisk som genetisk er bestemt for å modne som tert vokser bedre som en del av denne strategien. De har da generelt høyere appetitt eller fødemotivasjon enn MSW-laks, fordi de skal modne Kadri *et al.* (1996). På den andre siden er det fullt mulig å tenke at laksene med høyere vekst modner som tert fordi de har nådd en terskel på størrelse (kombinasjon av lengde og vekt), fettnivå eller vekstrate.

Hva som er riktig av disse to teoriene, lar seg vanskelig si ut fra dette forsøket. Vi har gjort en observasjons-studie av vekstmønsteret hos tert og MSW-laks. Vi kan derfor ikke så lett skille årsak og virkning. Begge teoriene kan godt være riktige, men i dette forsøket klarer vi ikke å skille faktorene fra hverandre (årsak og virkning).

Det er fire mindre heldige forhold/faktorer som vanskeliggjør tolkingen:

1. Lysgruppene «vokser sammen» i sjøfasen slik at den forskjellen vi hadde i størrelse som smolt minker utover høsten.
 - De ulike smoltstørrelsene i lysregimene som kunne resultert i ulikt antall modne gjør ikke det (i særlig grad).
 - På den annen side er dette en indikasjon på at beslutningen om å modne som tert taes i sjøvannsfasen.
2. Få modne laks, slik at vi får et spinkelt statistisk grunnlag.
 - Dette medførte også et sterkt ubalansert datasett, m.h.p. antall modne og umodne.
3. Vi kjenner ikke det genetiske grunnlaget, slik at dette blir en ukjent faktor
 - Det er kjent at ulike stammer har tendens til å modne ved ulik alder. Det samme gjelder familier.

- Dette medfører at vi kan ha en sammensetning av ulike familier med ulike strategier med hensyn på vekst og modning. Dette kan ha gitt et uklart bilde av «mekanismen» for å modne tidlig.
 - Vi kan ha en sammensetning av fisk med overvekt av familier som vokser hurtig og modner tidlig.
4. Det er ikke målt kjønnshormoner underveis, slik at det blir vanskeligere å skille årsak og virkning.
- Dersom gonadotropin(1) nivået hadde vært målt, (eller hormoner tidligere i syklusen) kunne vi tidfeste omtrentlig når akkurat disse fiskene i dette forsøket, starter modningen.

Et viktig vendepunkt i kjønnsmodningen for villaks er når laksen begynner å vandre tilbake til elven fra havet. Vandringshastigheten i havet har man begrenset kunnskap om. Man kan ikke si nøyaktig når de starter tilbakevandringen til elven. Men man kan beregne noenlunde når de seinest kan ha startet for å ha nådd fram til kysten og elven. Schandy og Røed (2002) angir en vandringshastighet på mellom 5 og 10 mil i døgnet. Selv om vår fisk er oppdrettslaks i et kar, kan villfiskens atferd likevel være relevant. Oppdrettslaksen følger på mange måter samme årssyklus som en villaks. Siden laksen ankommer elven og gyter over en relativt vid periode, er det likevel ikke så lett å tidfeste oppstart av kjønnsmodningen på denne måten.

De to først faktorene i listen overfor har med selve biologien å gjøre. At fisken under LD24:0-lysregime vokste dårlig i sjøen var «uheldig». Dette i den forstand at da hadde fisk som var ca. 50 prosent tyngre enn de andre gruppen. Dersom dette ga flere modne kunne man predikert at stor smolt gir også medfører flere modne (i den grad det ikke var lysregime medvirket).

At man fikk få modne kunne man kanskje predikert, fordi omtrent 10 prosent var en vanlige andel modne som tert innen oppdrett. Man måtte da hatt mange flere fisker i forsøket for å unngått dette. Men det kan ha hatt med ressursituasjonen å gjøre at man ikke brukte nok fisker. De to siste faktorene har mer med designet av forsøket eller med for «ambisiøse» hypoteser å gjøre. «Det er lettere å være etterpå klok».

11 Etterord

11.1 Sammendrag av resultater

11.1.1 Biologi

1. Gonader (organer med rogn eller melke) og GSI (se def. i kap 3.)
 - GSI og gonadevektene har en antydning til «tretoppet» fordeling, og ikke bimodal som mange angir. Dette er antydnet i figur 4.9 på side 50.
2. Postsmoltmodne=laks som gyter etter en sommer i sjø.
 - De postsmoltmodne er antageligvis en blanding av dverghanner (dvs. har ikke vært smolt) og ekte postsmoltmodne. Dette er vist i figur 4.1 på side 40.
3. LD24:0-lysregime dvs. 24 timer lys i ferskvannsfasen
 - a.) Påvirker ikke bare gjennomsnittsvekten og -lengden, men også fordelingen dvs. spredningen. Dette er vist i figur 6.9 på side 67.
 - b.) Påvirker k-faktoren, spesielt hunnene får en voldsom variasjon i forholdet mellom vekt og lengde. Dette er vist i figur 6.12 på side 70.
4. Parren dvs. liten laks i ferskvann
 - Parren har sannsynligvis en vekststrategi avhengig av hvordan den oppfatter årstiden, og den velger muligens en optimal smoltstørrelse. Dette er vist i figur 8.1 på side 123 og i figur 8.2 på side 126.
5. Generell vekst i sjøfasen
 - Veksten hos laks styres i sjøfasen av årstidsoppfattelse, og ikke bare av temperatur. Dette er grundig dokumentert i Forsberg (1995).
6. Kjønn

- Kjønn bestemmer i liten grad vekten i det første året i sjøen. Dette er vist i kapittel 9.
7. Tert=laks som gyter etter en vinter i sjø
- Modnende laks (tert) ser ut til å ha økt lengde og vekt i sjøvannfasen (i dette forsøket). Vist i kapittel 10.

Jeg oppfatter resultatene 1. og 3b. som ukjente eller relativt sjeldne (få har sett på dette). Resultatene 3a. og 7. er forhåpentligvis interessante.

11.1.2 Statistikk

1. ANCOVA

- ANCOVA-modellen må tilpasses med tredjegradsledd og variansfunksjoner for å fungere på dette datasettet. Dette er vist i kapittel 8.

2. MANCOVA

- ANCOVA-modellens tilpasninger som jeg har gjort kan få konsekvenser ved tilpasning av modell i MANCOVA.
- Dersom kovariatene ikke er like i gruppene kan det få konsekvenser i MANCOVA.

3. Valg mellom MANCOVA og MANOVA

- Vil få konsekvenser for hvilke hypoteser man i praksis tester, når man bruke en «indre» kovariat. Dette er diskutert i kapittel 7 og 10.

4. Lme-rutinen

- Lme-rutinen kan brukes som en variant av Diggles modell (slik jeg har gjort det med tredjegradspolynom og individeffekter).
- Utvidet bruk av individeffekter ser ut til å løse noe av problemet med relativt ulik vekst hos individene.
- Kan gi grei balanse mellom residualer og individeffekter. Det er ingen perfekt løsning, da residualene delvis kommer fra mistilpasninger av modellen (Kan likevel godt nok til å finne forskjeller mellom grupper av fisk). LME er diskutert i kapittel 7 og bruken vist i appendiks D.

Resultatene 1-4 kan være relevante ved bruk av metodene på lignende datasett. Resultat 4. kan også være relevant ved bruk av «ny måleteknologi».

11.2 Perspektiver

11.2.1 Datasettet

De to siste målingene

Jeg har ikke sett på de to siste målingene, men her kunne man arbeidet med en litt annen hypotese. Man kunne f.eks. undersøkt i hvilken grad terten stopper å vokse på det individuelle plan. eventuelt om dette kan predikere om et individ er en tert eller modner senere. Man kan også se om det er mulig å skille mellomlaks (2SW) fra storlaks (3SW), på grunnlag av GSI og vekstmønster.

Flere målinger i ferskvannsfasen

Som jeg har påpekt tidligere ser det ut til å være gjort tre målinger til i ferskvannsfasen. Det ville derfor være interessant å se på denne fasen, basert på de individuelle målingene med multivariate metoder.

SGR og kondisjonsfaktor

Jeg har ikke i denne oppgaven sett på SGR, uten i den grad størrelsen kommer til uttrykk i utviklingen over tid av lengde og vekt. Kondisjonsfaktor har jeg heller ikke sett noe videre på utenom at jeg påviser effekt av lysregime på kondisjonsfaktor. For eksempel har jeg ikke sett på forskjell i kondisjonsfaktor hos modne og umodne.

Det fins flere interessante resultater i materialet enn de jeg har valgt å ta opp i denne oppgaven.

11.2.2 Statistiske metoder

Flere metoder

Videre er det mulig å bruke logistisk regresjon, EM-algoritmen og diskriminantanalyse for å se om man kan få mer ut av materialet. Likeså kan man kombinere flere mål inn i en analyse, og ikke bare se på lengde og vekt separat slik jeg i hovedsak har gjort.

Simuleringer

Planen var opprinnelig å simulere data med lignende strukturer etter å ha kartlagt datasettet. Dette for å se på styrke og robusthet til de ulike metodene. Dette mener jeg ville vært det mest interessante å gjøre rent statistisk på grunnlag av datasettet. Å studere uteliggernes innflytelse i modellene, kunne også vært spennende.

Meta-analyse

Meta-analyse ser ut til å komme mer i bruk på flere felter (Scheiner og Gurevitch 2001), for eksempel innen medisin. Jeg mener dette også kunne ha noe for seg innen det biologiske fagområdene jeg arbeidet med i denne oppgaven. For å ta fatt på dette kreves det etter min oppfatning kunnskaper innen både biologi og statistikk.

Feilbruk av metoder

Leser man artikler som for eksempel Peterman (1990), Johnson (1999), Jennions og Møller (2003) og Hoenig og Heisey (2001) får man et innblikk i en omfattende misbruk og feiltolkning av statistiske metoder. Peterman (1990) peker på at ikke signifikans blir tolket til ikke effekter, mens Jennions og Møller (2003) finner lav styrke i mange undersøkelser innenfor biologi. Hoenig og Heisey (2001) viser at å regne ut styrken det i observerte alternativ ikke gir ytterligere informasjon, fordi p-verdien og styrken henger entydig sammen.

Mitt inntrykk er at resultatene som rapporteres er mer pålitelige enn det man kan få inntrykk av ved å lese de nevnte artikler. Grunnen til dette er at statistikken ikke brukes alene, men settes inn i en faglig sammenheng, og ofte er tuftet på god biologisk kunnskap.

Betydelige forskjeller i middelværdier har som regel en biologisk forklaring, og disse vil framtre enten man bruker multivariate metoder eller for eksempel ANOVA. Grafikk vil ofte være en støtte for de statistiske metodene. Bruk av grafikk er en vesentlig del av undersøkelsen av ett materiale. Det håper jeg å ha vektlagt sterkt nok i denne oppgaven.

11.3 Sluttord

11.3.1 Måloppnåelse

De opprinnelige målene med oppgaven var:

- Å finne enklest mulige statiske metoder til å bruke på datasettet som jeg har.
- Å implementere metodene i Splus.
- Å gjøre tilpasninger av modellene.
- Å evaluere metodene.

De tre første er rimelig bra oppfylt men det er ikke nummer 4. Det vil si evaluering av metodene. Man kan ikke evaluere metodene ut fra et enkelt datasett. Heller bør man simulere realistiske data for å sammenligne metodene skikkelig.

De justerte målene var:

- Å finne enklest mulige statistiske metoder til å bruke på lignende datasett som det jeg har dvs. med gjentatte målinger over tid på samme individ.
- Å se teoretisk på hvilke metoder som er aktuelle ved «ny måleteknologi» (se innledningen) fortrinnsvis modellen til Diggle (1990).

Disse to sistnevnte mål er i rimelig grad nådd etter min oppfatning. Som en bieffekt av oppgaven ser det også ut til å kunne bli utviklet bedre måleutstyr for fisk, slik at man slipper å gjøre dette manuelt i forskningsøyemed.

11.3.2 «Dataverden»

Under hovedfag i akvakultur fikk man kun gjøre ett søk i databasen ASFA. I kontrast til dette kan man idag søke så mye man vil i ulike databaser, på nettet og i elektroniske tidsskrifter. Det er idag mulig å nøste både bakover og framover i tid med utgangspunkt i en artikkel. Dette vil jeg betegne som en revolusjon i effektivitet dersom man vil sette seg inn i et emne. Etter å ha vært nesten ti år vekke fra «vitenskapen» så oppleves dette som en enorm forbedring.

Når det gjelder statistikk-programmer er det ikke min oppfatning at en udelt forbedring har funnet sted. Selve programmene har blitt mer omfattende med flere rutiner, men i noen tilfeller har manualene blitt dårligere eller gies ikke lenger ut. Her er det store forskjeller mellom det ulike programleverandører. Jeg vil derfor kun uttale meg om Statistica som jeg hadde kjennskap til for ti år siden, og Splus som jeg nå har stiftet bekjentskap med. Statistica trykker ikke lenger opp en manual, og istedenfor må man hente bruddstykker her og der fra en elektronisk manual.

For Splus er situasjonen en litt annen. Der finnes nærmere fem hundre bøker om hvordan programmet skal brukes innen ulike områder. Jeg våger meg likevel på en liten kritikk av programmets «hjelpfiler». En del etablerte statistiske begreper finnes ikke i hjelpefilen, for eksempel MANCOVA og profilanalyse som er en kontrast i MANOVA. Som et eksempel vil jeg også nevne hjelpesiden for MANOVA, som jeg slett ikke synes er fyldestgjørende. På den annen side er noen av bøkene som er skrevet om bruk av Splus glimrende. Her vil jeg fremheve Crawley (2002) som en på mange måter flott bok for anvendelse av statistikk. Denne er tilpasset et nivået rett over et grunnleggende statistikkurs.

For meg har det vært en større utfordring å få Splus til å gjøre det jeg vil, enn å sette seg inn i prinsippet bak de statistiske metodene. På den annen siden har jeg kanskje ikke i nok grad benyttet meg av relevante nettsted for få hjelp eller rettet forespørsler via nettet til kompetente fagfolk og diskusjonssider. «Kritikk slår lett tilbake».

Figurer

1.1	Illustrasjon av laksegård (fangstanlegg som stenger av elven), fra Gudenåen i Danmark. Man trodde lenge at laksen i denne elven var utryddet, men den er nå i ferd med å ta seg opp igjen. Dette var tidligere en elv med betydelige fangster. Modifisert etter Jensen (1982).	2
1.2	Individuelle plott (Vekt4-6). (Individer er valgt ved «random» funksjon). Modell: tredjegradspolynom med modning nøstet innen kjønn. (Se kap. 7 og 10 for flere opplysninger).	11
2.1	Illustrasjon av ulike stadier i livssyklusen til laksen . Relative størrelser er ikke riktige. Rogn (nederst til venstre), øyerogn, plommeseekkyngel, parr og smolt (øverst til høyre) og voksen laks (gjeldfisk, dvs. ikke ytre tegn på å være kjønnsmoden). Illustratør: Jan Henning Olsen.	18
2.2	Illustrasjon av parr (øverst) og smolt (nederst). Modifisert etter Hansen (1998).	19
2.3	Enkel modell av forholdet mellom faktorene i eksperimentet. Tre faktorer påvirker modningen, som igjen påvirker veksten. Pilene angir årsaksretning.	20
2.4	Illustrasjon av Peder Claussøn Friis (1545–1614), som skrev at laksen vender hjem til den elv den er født i. Han mente også at «Laksestørjen ledet laksen, men Gud ledet laksestørjen». Modifisert etter Berg (1986).	22
2.5	Mer kompleks modell av forholdet mellom faktorene i eksperimentet. Pilene angir årsaksretning. Det kunne også gått pil fra lys direkte til tertmodning, og fra vekstfase1 til tertmodning. Her gjelder modellen innen kjønn.	26
3.1	Tabell av måletidspunkt: Utfra Vestfossen (1991) og utkast til oppgave. . .	35
3.2	Tabell av måletidspunkt: Utfra Stefansson <i>et al.</i> (1991)	35
4.1	Histogram av vekt hos fisk som er klassifisert som modne første høsten i sjøen. De er klassifisert ut fra ytre kjennetegn.	40

4.2	Illustrasjon gonader hos ørret (hunner). Hos laks er gonadene ikke vesentlig forskjellig. Nummeringen har med utvikling utover i gytesesongen (1, 2 og 3 er mest illustrerende for tidlige stadier innen «gytesesongen»). Modifisert etter Sømme (1954).	42
4.3	Illustrasjon av GSI hos hanner (øverst) og hunner (nederst). Horisontal akse er antall dager etter 8 juli 1990. Jeg antar kurvene er fra utvalg av ulike fisk, og at «knekkpunktene» er noen av de reelle måletidspunkt. Modifisert etter Thorpe (1994).	44
4.4	Illustrasjon av gonadevekt mot GSI hos hanner, inkludert postsmoltmodne (i forsøket som behandles i denne oppgave).	45
4.5	Illustrasjon av gonadevekt mot GSI hos hunner, inkludert en postsmoltmoden (i forsøket som behandles i denne oppgave).	46
4.6	Illustrasjon av gonadevekt i gram (x-akse) mot GSI (y-akse) hos hanner, inkludert postsmoltmodne. Her har jeg «zoomet» gradvis mer inn i nederste hjørne til venstre. Linjen i midterste plott representerer en mulig grense.	48
4.7	Histogrammene er delt i hunner til venstre (f) og hanner til høyre (m): Øverst gonadevekt mot GSI (< 15 gram). Midten venstre GSI (< 1) og midten høyre GSI (< 1,2). Nederst GSI (> 1).	49
4.8	Illustrasjon lengde (venstre) og vekt (høyre) mot GSI (<1,2) øverst og mot gonadevekt (<15 g) nederst. Begge kjønn er inkludert.	50
4.9	Illustrasjon av lengde (venstre) og vekt (høyre) mot GSI. Hanner øverst og hunner venstre.	50
6.1	Histogrammer av lengde i mm og vekt i gram ved forsøksstart. Øverst lengde og nederst vekt. Ett histogram pr. lysbehandling (Gruppe 1 er SNL , gruppe 2 er SNL+B og gruppe 3 er LD24:0).	58
6.2	Histogrammer av kondisjonsfaktor ved forsøksstart(Øverst)og spredningsplott av lengde mot vekt (nederst). Delt i lysbehandling (Gruppe 1 er SNL , gruppe 2 er SNL+B og gruppe 3 er LD24:0).	59
6.3	Histogrammer av lengde i mm og vekt i gram ved forsøksstart. Øverst lengde og nederst vekt. Ett histogram pr. kjønn. Hanner (1) til venstre og hunner (2) til høyre.	61
6.4	Histogrammer av lengde i mm og vekt i gram ved forsøksstart. Øverst lengde og nederst vekt. Ett histogram pr. modningskategori (1=umoden, 2=tert, 3=postsmoltmodne)	62

6.5	Spredningsplott av vekten 26 mai. 1988 (vekt2) mot vekten 15. september 1987 (vekt1). Plottet er delt opp i tre etter lysbehandlingene. Symboler; 1=SNL : åpen sirkel, 2=SNL+B : kors og 3= LD24:0 : trekant.	63
6.6	Spredningsplott av startlengde mot lengde2 NB! X-akse i mm Y-akse i cm. Symboler som i forrige figur. Linjene representerer glatting med loess-rutinen (opsjoner: span=0.5, førstegrads og symmetrisk).	64
6.7	Spredningsplott av kondisjonsfaktor1 mot kondisjonsfaktor2. Et plott lysbehandling (1=SNL, 2=SNL+B og 3= LD24:0).	65
6.8	Spredningplott av vekt1 mot vekt2. Plottet er delt i hanner=1 og hunner=2. Ulike symboler etter lysbehandling (åpen sirkel=SNL, kors=SNL+B og trekant= LD24:0). Linjene representerer ikke-parametrisk regresjon ved loess.	66
6.9	Histogrammer av smolten (lengde i mm og vekt i gram). Øverst lengde og nederst vekt. Ett histogram pr. lysregime (1=SNL, 2=SNL+B og 3= LD24:0).	67
6.10	Spredningsplott av «lengde2» mot «vekt2». Et plott pr. lysbehandling (1=SNL, 2=SNL+B og 3= LD24:0).	68
6.11	Spredningsplott av «lengde2» mot «kondisjonsfaktor2». Et plottet pr. lysbehandling (1=SNL, 2=SNL+B og 3= LD24:0).	69
6.12	Spredningplott av vekt mot lengde. Ulike symboler etter lysbehandling. 1=SNL, 2=SNL+B og 3= LD24:0 (Øverst måling 1. og nederst måling 2.)	70
6.13	Spredningplott av vekt mot lengde. Ulike symboler etter lysbehandling. 1=SNL, 2=SNL+B og 3= LD24:0 (Øverst måling 3. og nederst måling 4.)	71
6.14	Spredningplott av vekt mot lengde. Ulike symboler etter lysbehandling. 1=SNL, 2=SNL+B og 3= LD24:0 (Øverst måling 5. og nederst måling 6.)	72
6.15	Spredningplott av vekt mot lengde. Ulike symboler etter lysbehandling. 1=SNL, 2=SNL+B og 3= LD24:0 (Øverst måling 7. og nederst måling 8.)	73
6.16	Spredningplott av «vekt2» mot «vekt3». Ulike symboler etter lysbehandling (1=SNL, 2=SNL+B og 3=LD24:0).	74
6.17	Spredningsplott av vekt3 mot vekt2 . Linjene representerer loess (). Plottet er delt opp i tre, etter lysbehandling. 1=SNL, 2=SNL+B og 3= LD24:0.	75
6.18	Histogram av vekt3. Histogrammet er delt opp i tre, etter lysbehandling. 1=SNL, 2=SNL+B og 3= LD24:0.	76
6.19	Plott av vekt-utviklingen pr. individ. Plottet er delt opp i seks, etter kjønn (1= hanner og 2=hunner) og lysbehandling (1=SNL, 2=SNL+B og 3= LD24:0).	77
6.20	Plott av en kurve for hver enkelt fisk (umodne hanner; nederst til venstre, umodne hunner; nederst til høyre, modne hanner; øverst til venstre og modne hunner; øverst til høyre).	78

8.1	Spredningsplott av vekten 26. mai 1988 (vekt2) mot vekten (minus middelvekten til alle grupper) 15. september 1987 (vektsentret1). Linjene representerer modellen. Plottet er delt opp i tre, etter lysbehandling. 1 = SNL, 2 = SNL+B og 3 = LD24:0.	123
8.2	Spredningsplott av lengden 26. mai 1988 (lengde2) mot startlengde (minus middelstartlengden til alle grupper) Linjene representerer modellen. Plottet er delt opp i tre, etter lysbehandling. 1 = SNL, 2 = SNL+B og 3 = LD24:0.	126
8.3	Spredningsplott av lengde2 mot vekt2. Linjene representerer modellen. Plottet er delt opp i tre, etter lysbehandling. 1 = SNL, 2 = SNL+B og 3 = LD24:0.	131
8.4	Spredningsplott av loglengde2 mot logvekt2. Linjene representerer modellen. Plottet er delt opp i tre, etter lysbehandling. 1=SNL, 2=SNL+B og 3=LD24:0.	132
A.1	Diagnostiske plot for modell: vekt2 forklart med tredjegradspolynom. Forklaringvariabel er vekt1.	175
A.2	Diagnostiske plot for modell: lengde2 forklart med tredjegradspolynom. Forklaringvariabel er lengde1.	176
A.3	Diagnostiske plot for modell for forholdet mellom lengde2 og vekt2. Modellen antar lik potens (tredje), men ett vertikalt løft avhengig av lysregime.	177
B.1	Spredningsplott av residualer. Plottet tilsvare korrelasjonen under diagonalen: vekt3-6: Modell: lys i samspill med kjønn. Tilhørende korrelasjonsmatrise er gitt i tabell B.4 på side 182 og kovariansmatrisen er gitt i tabell B.5 på side 183.	182
C.1	Korrelasjonsplott av residualer (Vekt3-6). Plottet tilsvare korrelasjonen under diagonalen. Modell: modning nøstet innen kjønn.	184
C.2	Diagnostiske plott: Manova: modning nøstet innen kjønn. Øverst histogram residualer vekt3-6. Nederst normalplott for vekt3-6.	186
D.1	Plott av modell for fixed effekt mot tid (mnd, umodne hanner) (øverst), og plott residualer mot tid (dager, begge kjønn) (nederst).	190

D.2	Plott av modell for fixed effekt mot tid. Modell: tredjegradspolynom i samspill med modning nøstet innen kjønn. Fordelingen av data på hvert tidspunkt er vist med punkter og linjer som går gjennom midpunktene i histogrammer.	192
D.3	Plott av residualer mot tid. Modell: tredjegradspolynom i samspill med modning nøstet innen kjønn (hanner=1, hunner=2, umoden=1 og moden=2).	193
D.4	Individuelle plott (Individer valgt ved «random» funksjon). Modell: tredjegrads modell med modning nøstet innen kjønn (Vekt4-6). Komplette serier (lengde og vekt).	194
D.5	Histogram av random effektene (1 og 2). Modell: Vektutviklingen forklart med tredjegradspolynom i samspill med modning nøstet innen kjønn. . .	195
D.6	Histogram av random effektene (3 og 4). Modell: Vektutviklingen forklart med tredjegradspolynom i samspill med modning nøstet innen kjønn. . .	196

Tabeller

4.1	Grenser mellom modne og umodne hanner i utvalgte arbeider. «Vekt fisk» er gjennomsnittsvæker oppgitt i publikasjonene eller estimert fra «figurer».	43
4.2	Grenser mellom modne og umodne hunner i utvalgte arbeider. «Vekt fisk» er gjennomsnittsvæker oppgitt i publikasjonene eller estimert fra «figurer».	43
4.3	Antall postsmoltmodne (PSM), tert, multimodne (PSM og tert) og MSW. «Ukjent» er fisk som er gått ut av forsøket for terminering (unntatt «postsmoltmodne») Hanner og hunner slått sammen (fra tidligere utkast til oppgave i akvakultur). Rekkesommene fremkommer ved å ikke ta med tallene i parentes (NB! en fisk mangler). Forklaring av forkortelser er i første avsnitt i dette kapittel.	46
4.4	Antall tert og MSW. Hanner og hunner hver for seg.	46
6.1	Gjennomsnittlig startvekt, –lengde og – kondisjonsfaktor. Lys1 er SNL, lys2 er SNL+B og lys 3 er LD24:0.	60
7.1	Variansfunksjoner i gls-, lme- og nmle-rutinen, der l står for lys og i for individ.	95
7.2	Korrelasjonstrukturer i lme- og nmle-rutinen.	108
7.3	Strukturer kovariansmatrisen til random effektene i lme- og nmle-rutinen.	108
8.1	Koeffisienter for vekt2: tredjegradspolynom.	124
8.2	Konfidensintervall (95 prosent) for effekter, varians og residual. Modell: tredjegradspolynom for vekt2.	124
8.3	Koeffisienter for lengde2: tredjegradspolynom.	127
8.4	Konfidensintervall (95 prosent) for effekter, varians og residual. Modell: tredjegradspolynom for lengde2.	127
8.5	Koeffisienter: vekt forklart ved lengde i tredje + konstantledd.	129
8.6	Konfidensintervall (95 prosent) for effekter, varians og residual. Modell: vekt2 forklart ved $I(len2^3)$	130

8.7	Koeffisienter: logvekt forklart ved loglengde + konstantledd.	130
8.8	Koeffisienter: logvekt forklart ved loglengde + konstantledd.	131
8.9	Konfidensintervall (95 prosent) for effekter, varians og residual. Modell: logvekt2 forklart ved loglengde2.	132
9.1	Koeffisienter: ANCOVA-modell: tredjegradspolynom, vekt2 forklart med lys og kjønn.	138
9.2	Testobservatorer og tilhørende p-verdier for vekt3-6 (f-verdi er en tilnær- ming). MANOVA-modell: lys + kjønn (data= smolt.manova.alle4, con- trasts = list(kjønn = contr.treatment).	140
9.3	Koeffisienter: vekt 4–6: MANOVA-modell: Forklaringsvariabler: lys+kjønn.	140
9.4	Testobservatorer og tilhørende p-verdier for vekt3-6 (f-verdi er en tilnærming). MANCOVA-modell: vekt1 + lys + kjønn. (data= smolt.manova.alle4, contrasts = list(kjønn = contr.treatment).	140
9.5	Koeffisienter: vekt 4–6: MANCOVA-modell: Forklaringsvariabler: vekt1+lys+kjønn.	141
10.1	Fordeling av modne og umodne hanner og hunner.	144
10.2	Testobservatorer og tilhørende p-verdier for vekt3-6 (f-verdi er en tilnær- ming). MANOVA-modell: kjønn*modn.	145
10.3	Effekter på vekt. MANOVA-modell: kjønn i samspill med modning. . . .	145
10.4	Testobservatorer og tilhørende p-verdier for vekt3-6 (f-verdi er en tilnær- ming). MANOVA-modell: kjønn/modn.	146
10.5	Effekter i MANOVA-modell med modning nøstet innen kjønn.	146
10.6	Testobservatorer og tilhørende p-verdier for «elimineringmetoden» vekt3-6. MANOVA-modell: kjønn/modn.	148
10.7	Testobservatorer og tilhørende p-verdier for lengdet3-6 (f-verdi er en til- nærming). MANOVA-modell: kjønn/modn.	149
10.8	Effekter på lengde3-6 i MANOVA-modell med modning nøstet innen kjønn.	149
10.9	Testobservatorer og tilhørende p-verdier for «elimineringmetoden» lengde3-6. MANOVA-modell: kjønn/modn.	149
10.10	Testobservatorer og tilhørende p-verdier for vekt3-6 (f-verdi er en tilnærming). MANCOVA-modell: vekt1 + kjønn/modn. (data= smolt.manova.alle4, contrasts = list(modn = contr.treatment).	150
10.11	Effekter i MANCOVA-modell med modning nøstet innen kjønn. Startvekt som kovariat.	151

10.12	Testobservatorer og tilhørende p-verdier for vekt3-6 (f-verdi er en tilnærming). MANCOVA-modell: vekt2 + kjønn/modn. (data= smolt.manova.alle4, contrasts = list(modn = contr.treatment)).	151
10.13	Effekter i MANCOVA-modell med modning nøstet innen kjønn. Smoltvekt som kovariat.	151
10.14	Testobservatorer og tilhørende p-verdier for lengde3-6 (f-verdi er en tilnærming).MANCOVA-modell: len1 + kjønn/modn. (data= smolt.manova.alle4, contrasts = list(modn = contr.treatment)).	152
10.15	Effekter i MANCOVA-modell med modning nøstet innen kjønn. Lengde1 som kovariat.	152
B.1	Koeffisienter: ANCOVA-modell:tredjegradspolynom vekt2 forklart med lys. (Samme data som med kjønn)	179
B.2	Koeffisienter: Modell:tredjegradspolynom vekt med lys og kjønn og ulike varianser for kombinasjoner av lys og kjønn.	180
B.3	Koeffisienter: manova vekt4-6:lys*kjønn.	181
B.4	Korrelasjonmatrise for vekt3-6; modell lys i samspill med kjønn.	182
B.5	Kovariansmatrise for vekt3-6; modell lys i samspill med kjønn.	183
C.1	Korrelasjonmatrise for vekt; modell med modning nøstet innen kjønn. . .	185
C.2	Kovariansmatrise for modell med modning nøstet innen kjønn.	185
C.3	Korrelasjonmatrise for lengde; modning nøstet innen kjønn.	185
C.4	Kovariansmatrise for lengde; modning nøstet innen kjønn.	187
C.5	Kovariansmatrise for lengde; Modell: MANCOVA med startlengde som covariat, og modning nøstet innen kjønn.	187
C.6	Korrelasjonmatrise for lengde; Modell: MANCOVA med startlengde som covariat, og modning nøstet innen kjønn.	188
D.1	Effekter for vekt i lme-modell: modning nøstet innen kjønn. Lineær modell: antall dager fra dag 366 som forklaringsvariabel.	191

A Lys

A.1 Ferskvannsfase

Splus utskrift for modell; vekt2 forklart med tredjegradspolynom;

```
Generalized least squares fit by maximum likelihood

Model: vekt2 ~ (lys * vektsentrert1) + I(vektsentrert1^2)
          +I(vektsentrert1^3)

          AIC      BIC    logLik
6546.067 6596.192 -3262.034

Variance function:
Structure: Different standard deviations per stratum
Formula:   ~ 1 | lys
Parameter estimates:
1          2          3
1          1.05      1.69

Residual standard error: 20.75
Degrees of freedom: 704 total; 696 residual
```

Splus utskrift for modell; lengde2 forklart med tredjegradspolynom;

```
Generalized least squares fit by maximum likelihood

Model: len2 ~ (lys * len1cmsentrert) + I(len1cmsentrert^2)
          + I(len1cmsentrert^3)
```

AIC	BIC	logLik
2599.013	2648.054	-1288.506

Variance function:

Structure: Different standard deviations per stratum

Formula: $\sim 1 \mid \text{lys}$

Parameter estimates:

1	2	3
1	0.96	1.14

Residual standard error: 1.77

Degrees of freedom: 638 total; 630 residual

Splus utskrift for modell; vekt2 forklart med lengde2 fastlåst i tredjepotens;

Generalized least squares fit by maximum likelihood

Model: $\text{vekt2} \sim \text{lys} + \text{I}(\text{len2}^3)$

AIC	BIC	logLik
5405.784	5437.51	-2695.892

Variance function:

Structure: Different standard deviations per stratum

Formula: $\sim 1 \mid \text{lys}$

Parameter estimates:

1	2	3
1	1.00	5.27

Residual standard error: 7.5

Degrees of freedom: 687 total; 683 residual

Splus utskrift for modell; regresjon: lnvekt2 forklart med lnlenge2;

Generalized least squares fit by maximum likelihood

Model: $\log(\text{vekt2}) \sim \text{lys} * \log(\text{len2})$

Data : smolt3

AIC	BIC	logLik
-792,6	-751,8	405,2

Variance function :

Structure : Different standard deviations per stratum

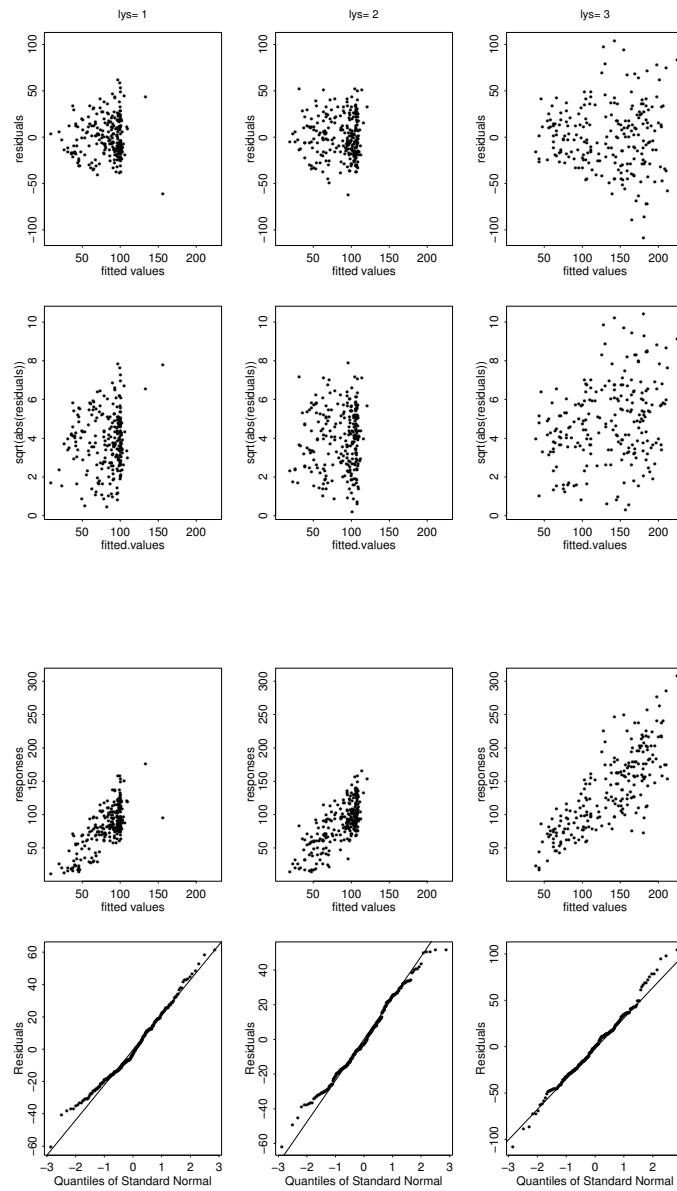
Formula : $\sim 1 \mid \text{lys}$

Parameter estimates :

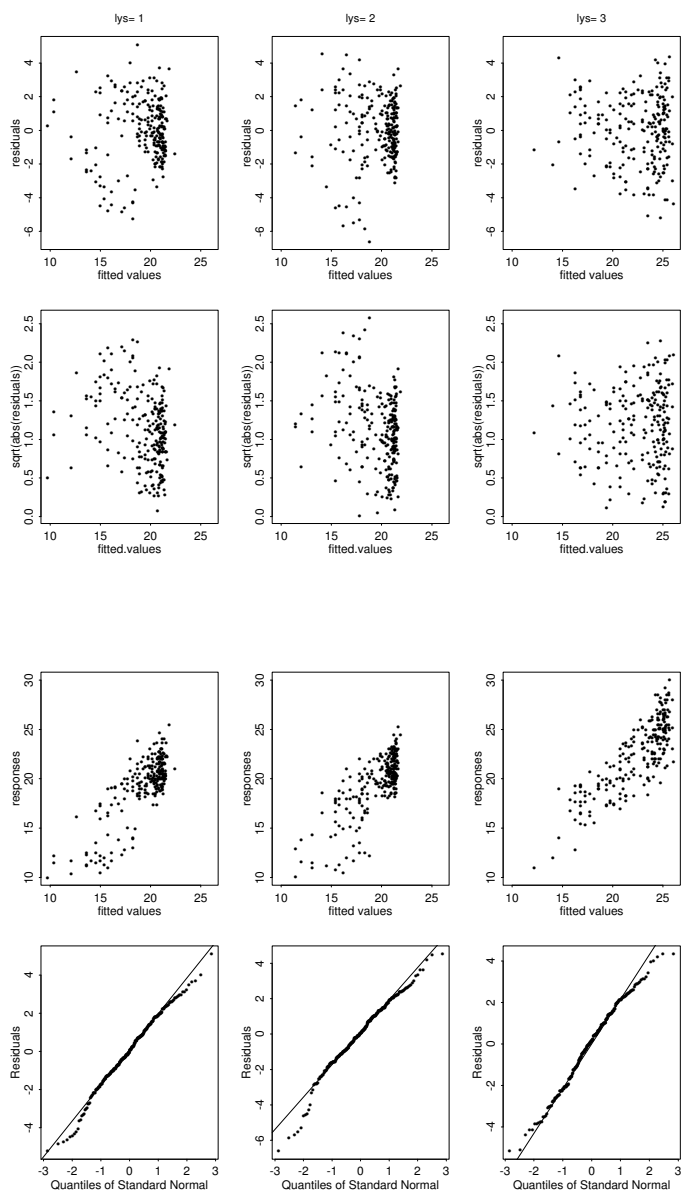
1	2	3
1	1,02	2,9

Residual standard error : 0,097

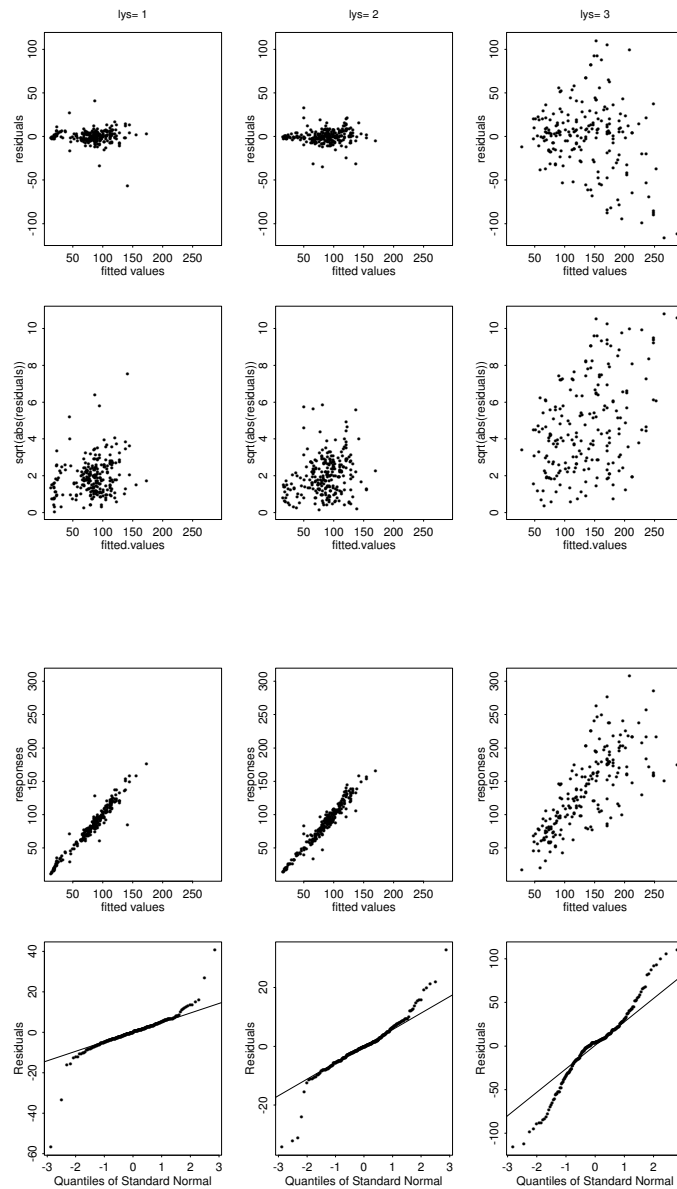
Degrees of freedom : 687 total ; 681 residual



Figur A.1: Diagnostiske plot for modell: vekt2 forklart med tredjegradsynom. Forklaringvariabel er vekt1.



Figur A.2: Diagnostiske plot for modell: lengde2 forklart med tredjegradspolynom. Forklaringvariabel er lengde1.



Figur A.3: Diagnostiske plot for modell for forholdet mellom lengde2 og vekt2. Modellen antar lik potens (tredje), men ett vertikalt løft avhengig av lysregime.

B Kjønn

Splus-utskriften med modell som i ferksvannsfasen (tredjegradspolynom og varIdent), bare med kjønn som ekstra faktor (uten samspill). Resultatene er gitt i tabell 9.1 på side 138.

```
Generalized least squares fit by maximum likelihood
```

```
Model: vekt2 ~ kjønn + (lys * vektsentrert1)
+ I(vektsentrert1^2) + I(vektsentrert1 ^ 3)
```

```
          AIC          BIC      logLik
4259.429 4309.004 -2117.714
```

```
Variance function:
```

```
Structure: Different standard deviations per stratum
```

```
Formula: ~ 1 | lys
```

```
Parameter estimates:
```

```
1      2      3
1      1.04    1.6
```

```
Residual standard error: 20.5
```

```
Degrees of freedom: 460 total; 451 residual
```

I Splus-utskift nedenfor har jeg brukt samme dataene som i modellen ovenfor men uten kjønn som faktor. Resultatene er også gitt i tabell B.1 på neste side.

```
Generalized least squares fit by maximum likelihood
```

```
Model: vekt2 ~ (lys * vektsentrert1) + I(vektsentrert1^2)
+ I(vektsentrert1^3)
```

```
Data: smolt3ok
      AIC      BIC    logLik
4257.433 4302.877 -2117.717
```

Variance function:

Structure: Different standard deviations per stratum

Formula: ~ 1 | lys

Parameter estimates:

```
1      2      3
1      1.0    1.6
```

Residual standard error: 20.5

Degrees of freedom: 460 total; 452 residual

Tabell B.1: Koeffisienter: ANCOVA-
modell:tredjegradspolynom vekt2 forklart med
lys. (Samme data som med kjønn)

sex3V2	verdi	st.feil	t-verdi	p-verdi
(Intercept)	94.5	2.1	44.8	< .0001
lys2	0.82	2.5	0.33	0.7415
lys3	53.9	3.3	16.4	< .0001
vektsentrert1	2.84	0.5	6.09	< .0001
$I(\text{vektsentrert1}^2)$	-0.47	0.069	-6.82	< .0001
$I(\text{vektsentrert1}^3)$	0.024	0.0058	4.06	0.0001
lys2vektsentrert1	0.98	0.51	1.90	0.0577
lys3vektsentrert1	6.2	0.73	8.51	< .0001

I Splus-utskriften nedenfor er resultater fra kjøring av modell med kjønn men med ulike varianser for alle kombinasjoner av lys og kjønn. Resultatene er også gitt i tabell B.2 på neste side.

```
Generalized least squares fit by maximum likelihood
  Model: vekt2 ~ kjønn + (lys * vektsentrert1)
+ I(vektsentrert1^2) + I(vektsentrert1^3)
```

Data : smolt3ok

AIC	BIC	logLik
4264.152	4326.121	-2117.076

Variance function :

Structure : Different standard deviations per stratum

Formula : ~ 1 | lys * kjønn

Parameter estimates :

1*1	1*2	2*1	2*2	3*1	3*2
1	0.9	1.0	0.9	1.5	1.5

Residual standard error : 21.9

Degrees of freedom : 460 total ; 451 residual

Tabell B.2: Koeffisienter: Modell:tredjegradspolynom vekt med lys og kjønn og ulike varianser for kombinasjoner av lys og kjønn.

sex4V2	verdi	st.feil	t-verdi	p-verdi
(Intercept)	94.24	2.56	36.8	< .0001
kjønn	0.19	2.25	0.08	0.9331
lys2	1.0	2.47	0.42	0.6739
lys3	54.1	3.29	16.4	< .0001
vektsentrert1	2.82	0.47	6.0	< .0001
$I(\text{vektsentrert1}^2)$	-0.47	0.069	-6.84	< .0001
$I(\text{vektsentrert1}^3)$	0.0245	0.00597	4.1	< .0001
lys2vektsentrert1	1.01	0.51	1.98	0.0484
lys3vektsentrert1	6.23	0.73	8.5	< .0001

I Splus-utskriften nedenfor er resultater fra MANOVA modell med samspill mellom lys og kjønn. Resultatene er også gitt i tabell B.3 på neste side.

Tabell B.3: Koeffisienter: manova vekt4-6:lys*kjønn.

ancovaVS	vekt3	vekt4	vekt5	vekt6
(Intercept)	404,47	986,88	1218,51	1633,36
lys1	-12,58	-3,48	8,10	46,20
lys2	21,09	6,03	-8,61	-22,92
kjonn	2,40	14,15	14,13	-24,76
lys1kjonn	2,93	-1,81	-6,89	-63,68
lys2kjonn	-5,50	4,56	12,37	-2,82

```
manova(formula = cbind(vekt3 , vekt4 , vekt5 , vekt6) ~ lys * kjonn
,
data = smolt.manova.alle4 ,
contrasts = list(kjonn = contr.treatment) , na.action = na.
exclude)
```

Terms:

	lys	kjonn	lys:kjonn	Residuals
Deg. of Freedom	2	1	2	364

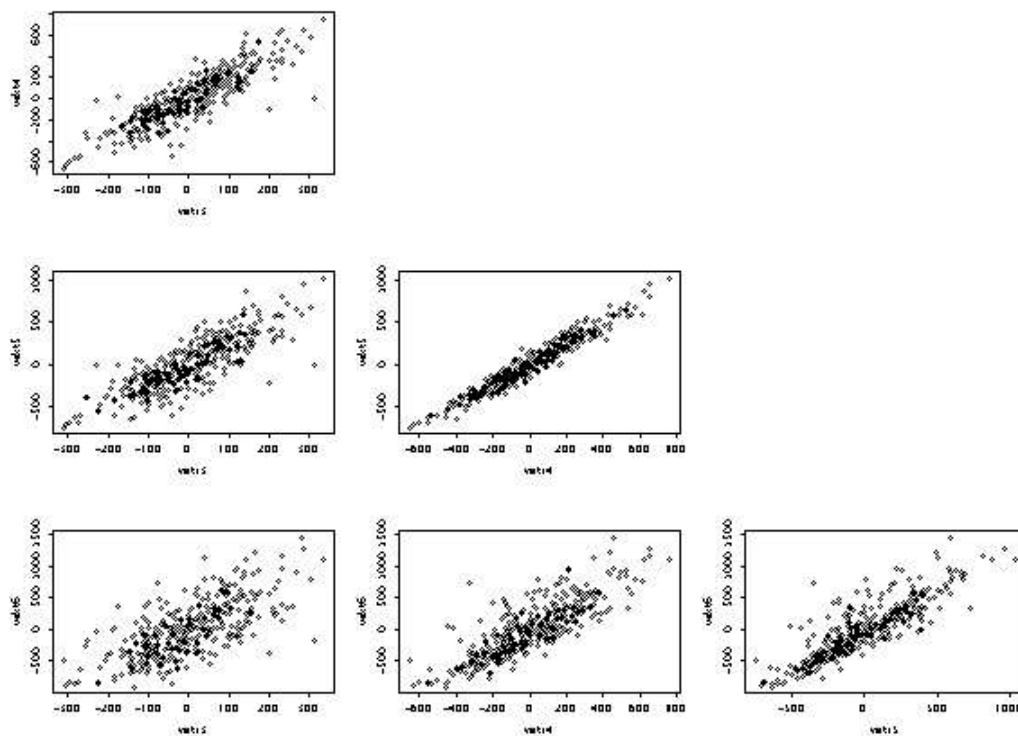
Estimated effects may be unbalanced

Analysis of Variance Table:

	Df	Pillai	Trace	approx. F	num df	den df	P-value
lys	2	0,157	7,726		8	724	0
kjonn	1	0,01	0,884		4	361	0,474
lys:kjonn	2	0,024	1,091		8	724	0,367

Splus-utskrift for MANOVA modell uten samspill mellom lys og kjønn er gitt nedenfor. Resultatene er også vist i tabell 9.2 på side 140 og tabell 9.3 på side 140.

```
manova(formula = cbind(vekt3 , vekt4 , vekt5 , vekt6) ~ lys +
kjonn , data =
```



Figur B.1: Spredningsplott av residualer. Plottet tilsvare korrelasjonen under diagonalen: vekt3-6: Modell: lys i samspill med kjønn. Tilhørende korrelasjonmatrise er gitt i tabell B.4 og kovariansmatrisen er gitt i tabell B.5 på neste side.

Tabell B.4: Korrelasjonmatrise for vekt3-6; modell lys i samspill med kjønn.

Korrelasjoner	vekt3	vekt4	vekt5	vekt6
vekt3	1,000	—	—	—
vekt4	0,866	1,000	—	—
vekt5	0,824	0,970	1,000	—
vekt6	0,692	0,825	0,859	1,000

Tabell B.5: Kovariansmatrise for vekt3-6; modell lys i samspill med kjønn.

Kovarians	vekt3	vekt4	vekt5	vekt6
vekt3	13326	–	–	–
vekt4	24594	60492	–	–
vekt5	29454	73954	95993	–
vekt6	35188	89371	117216	193898

```
smolt.manova.alle4 , contrasts = list(kjonn = contr.
  treatment) ,
na.action = na.exclude)
```

Terms:

	lys	kjonn	Residuals
Deg. of Freedom	2	1	366

Estimated effects may be unbalanced

Analysis of Variance Table:

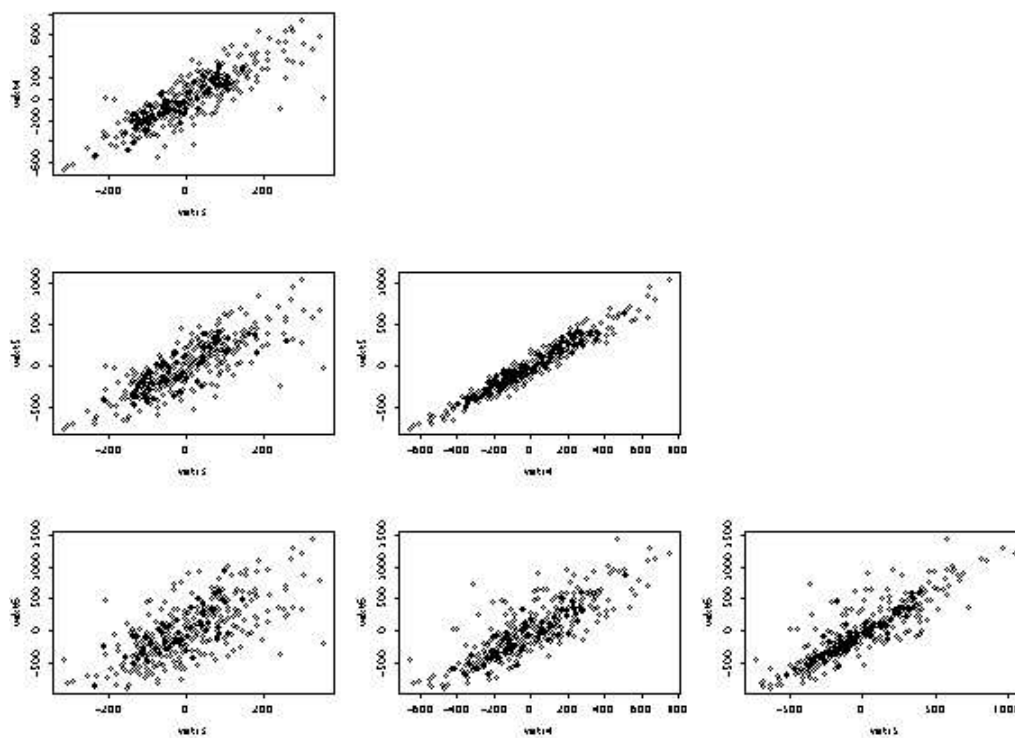
	Df	Pillai	Trace	approx. F	num df	den df	P-value
lys	2	0,156		7,7	8	728	0
kjonn	1	0,01		0,879	4	363	0,477
Residuals	366						

C Modning

C.1 MANOVA

C.1.1 Vekt

Her vises spredningsplot av residualene for MANOVA-modellen i figur C.1.1, samt tilhørende korrelasjonsmatrise og kovariansmatrise i tabell C.1 på neste side og i tabell C.2 på neste side.



Figur C.1: Korrelasjonsplott av residualer (Vekt3-6). Plottet tilsvarer korrelasjonen under diagonalen. Modell: modning nøstet innen kjønn.

Tabell C.1: Korrelasjonmatrise for vekt; modell med modning nøstet innen kjønn.

Korrelasjoner	vekt3	vekt4	vekt5	vekt6
vekt3	1,000	–	–	–
vekt4	0,855	1,000	–	–
vekt5	0,802	0,969	1,000	–
vekt6	0,675	0,830	0,864	1,000

Tabell C.2: Kovariansmatrise for modell med modning nøstet innen kjønn.

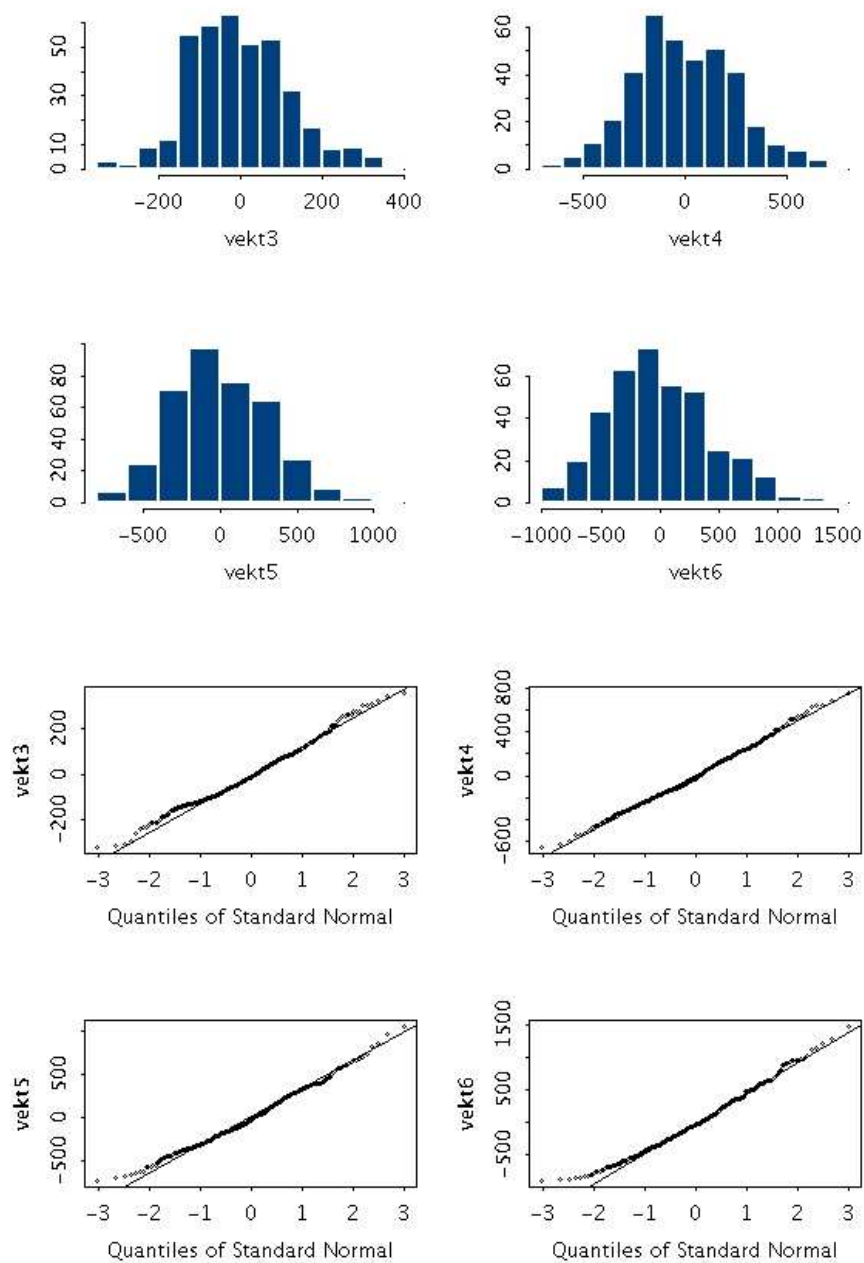
KovariansNV	vekt3	vekt4	vekt5	vekt6
vekt3	13960	–	–	–
vekt4	24856	60551	–	–
vekt5	29355	73915	96028	–
vekt6	34798	89193	116920	190516

Tabell C.3: Korrelasjonmatrise for lengde; modning nøstet innen kjønn.

Korrelasjoner	len3	len4	len5	len6
len3	1,000	–	–	–
len4	0,894	1,000	–	–
len5	0,855	0,967	1,000	–
len6	0,725	0,847	0,871	1,000

C.1.2 Lengde

Korrelasjonsmatrisen og kovariansmatrisen for MANOVA-modell er vist i tabell C.3 og i tabell C.4 på side 187.



Figur C.2: Diagnostiske plott: Manova: modning nøstet innen kjønn. Øverst histogram residualer vekt3-6. Nederst normalplott for vekt3-6.

Tabell C.4: Kovariansmatrise for lengde; modning
nøstet innen kjønn.

Kovariansmatrise	len3	len4	len5	len6
len3	9,91	–	–	–
len4	9,87	12,30	–	–
len5	10,18	12,82	14,3	–
len6	9,52	12,39	13,7	17,38

Tabell C.5: Kovariansmatrise for lengde; Modell:
MANCOVA med startlengde som covariat, og
modning nøstet innen kjønn.

Kovariansmatrise	len3	len4	len5	len6
len3	7,45	–	–	–
len4	7,11	9,20	–	–
len5	7,43	9,73	11,23	–
len6	6,69	9,22	10,57	14,13

C.2 MANCOVA

C.2.1 Lengde

Korrelasjonsmatrisen og kovariansmatrisen for MANOVA-modell er vist i tabell C.6 på neste side og i tabell C.5. Vi ser at variansen har minket i forhold til MANOVA-modellen.

Tabell C.6: Korrelasjonmatrise for lengde; Modell: MANCOVA med startlengde som covariat, og modning nøstet innen kjønn.

Korrelasjoner	len3	len4	len5	len6
len3	1,000	–	–	–
len4	0,858	1,000	–	–
len5	0,812	0,958	1,000	–
len6	0,652	0,808	0,839	1,000

D LME

Jeg har brukt et tredjegradspolynom der tiden er forklaringsvariabel. Modellen er beskrevet i 7.5. Verken et lineær modell eller et andregradspolynom er egnet. Jeg demonstrere likevel først en enkel linear modell for å vise at det ikke er godt nok. Jeg har også her snevret inn antall variabler jeg undersøker til måling nr. 3-6 (dvs mellom september 1988 og mai 1989). Jeg bruker modning nøstet innen kjønn.

D.1 Lineær modell

D.1.1 Vekt

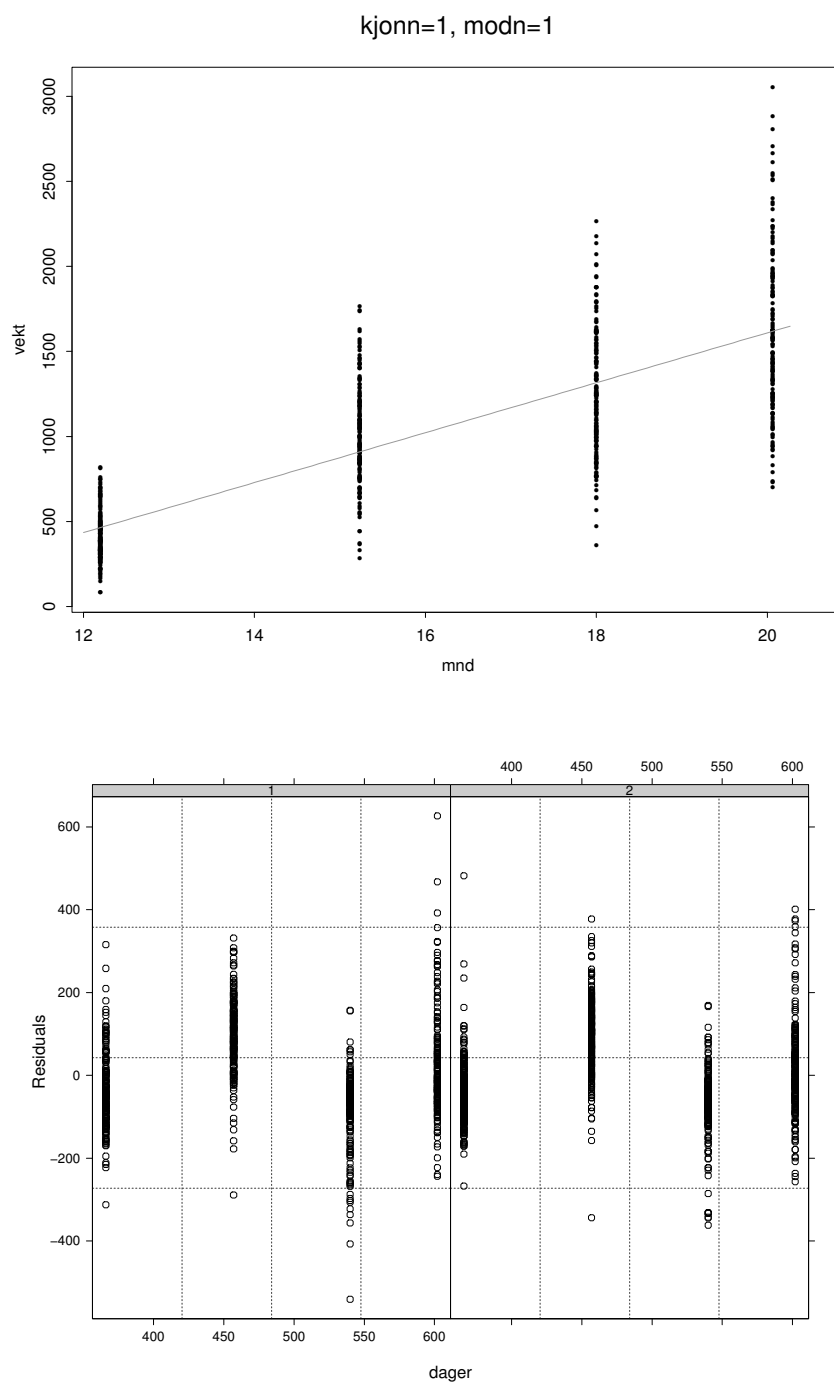
Uttrykker vi modellen uten lineær algebra blir den faste del av modellen (for den kontrast jeg kjører):

$$W_{kmi}(t) = \mu + \alpha_k + \alpha_{km} + \beta_1 t + \beta_{2k} t + \beta_{3km} t + \epsilon_{kmi} \quad (\text{D.1})$$

der t er tiden, μ er startvekt, α_k tillegg for kjønn, α_{km} tillegg for modning innen kjønn, β_1 er stigningstallet, $\beta_k t$ er tillegg i stigningstallet for kjønn og β_{km} er tillegg i stigningstallet for modning innen kjønn. (I tillegg kommer individ effektene.)

Jeg fikk allerede her resultater i harmoni med resultatene fra MANOVA/MANCOVA. I tabell D.1 på side 191 ser vi en effekt på stigningstallet for modne hanner (kjønn1modnI(dager - 366)) kontra umodne hanner. Figur D.1 på neste side) viser at modellen ikke er god. Dette fordi modellen ikke går sentralt gjennom dataene, og residualene opptrer systematisk i tid.

Jeg ser derfor ikke verdi i å kommentere modellen ytterligere. Andregrads polynom har ingenting for seg fordi vi da får enda mer mistilpasning.



Figur D.1: Plott av modell for fixed effekt mot tid (mnd, umodne hanner) (øverst), og plott residualer mot tid (dager, begge kjønn) (nederst).

Tabell D.1: Effekter for vekt i lme-modell: modning nøstet innen kjønn. Lineær modell: antall dager fra dag 366 som forklaringsvariabel.

lme1	Effekt	st.avik	frihetsgrader	t-verdi	p-verdi
(Intercept)	464.8	11.9	1320	39.0	< .0001
I(dager - 366)	4.9	0.1	1320	40.8	< .0001
kjønn	-26.9	15.4	489	-1.8	0.08
kjonn1modn	-82.5	29.4	489	-2.8	0.005
kjonn1modn	-77.7	41.7	489	-1.9	0.063
I(dager - 366):kjonn	-0.2	0.2	1320	-1.0	0.32
kjonn1modnI(dager - 366)	1	0.3	1320	3.3	0.001
kjonn2modnI(dager - 366)	0.5	0.4	1320	1.2	0.24

D.2 Modell med tredjegradspolynom

Siden jeg ikke gjør bruk av LME-modellen i den biologiske diskusjonen, viser jeg her kun en grafisk dokumentasjon av hvordan modellen virker.

Modellen er et tredjegradspolynom med individeffekter på alle leddene. Videre er modning nøstet innen kjønn. Vektutviklingen er forklart med tredjegradspolynom i samspill med modning nøstet innen kjønn.

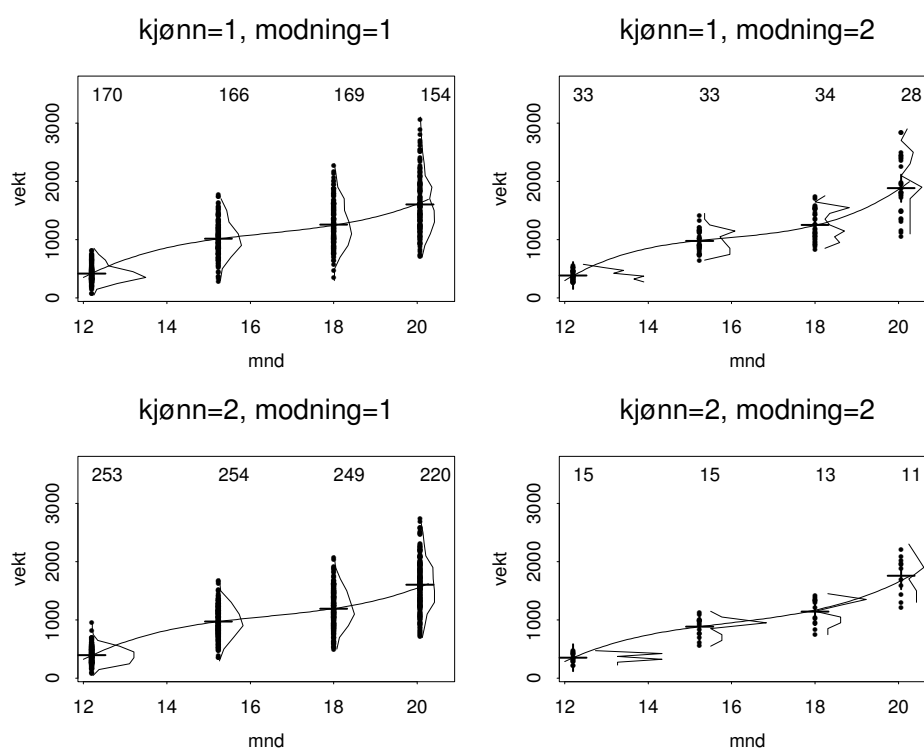
D.2.1 Vekt

Figur D.2 på neste side viser at modellen treffer bra sentralt i fordelingene.

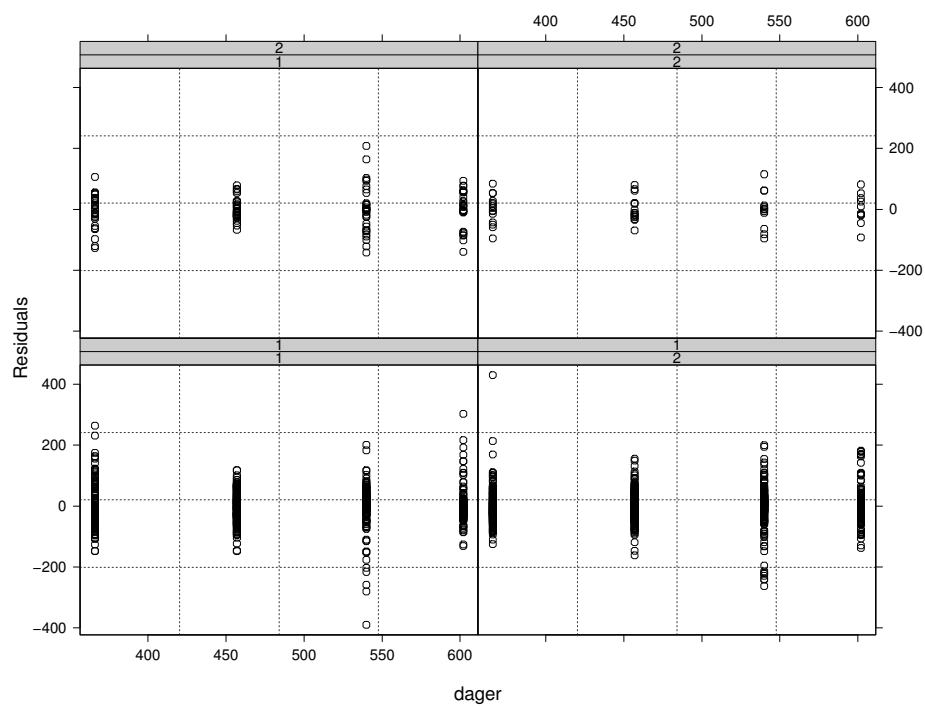
Residualplottet (figur D.3 på side 193) viser at residualene er relativt små på siste tidspunkt. Dette har med fleksibiliteten til individeffektene å gjøre.

Histogram av individ effektene (1- 4) er vist i figur figur D.5 på side 195 og figur D.6 på side 196. Vi ser at individeffektene er en del avvikende fra normalfordeling.

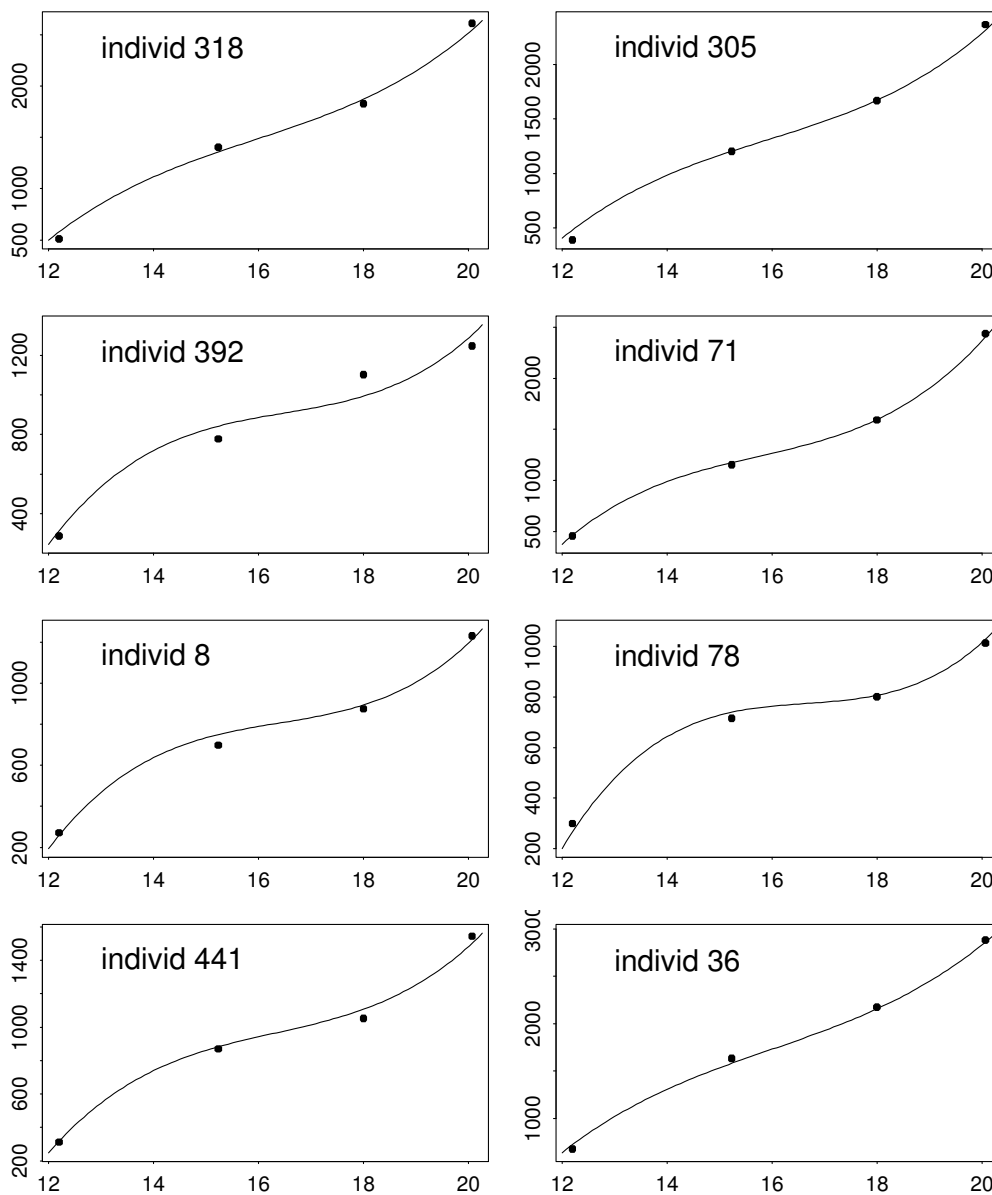
Sammenstill vi dette med histogrammet av individeffektene, får vi et forenklet bilde av modellen. Vi får det mest korrekte bildet ved å plote den predikerte linjen for hvert individ mot observasjonene for hvert individ. Dette er gjort i figur D.4 på side 194.



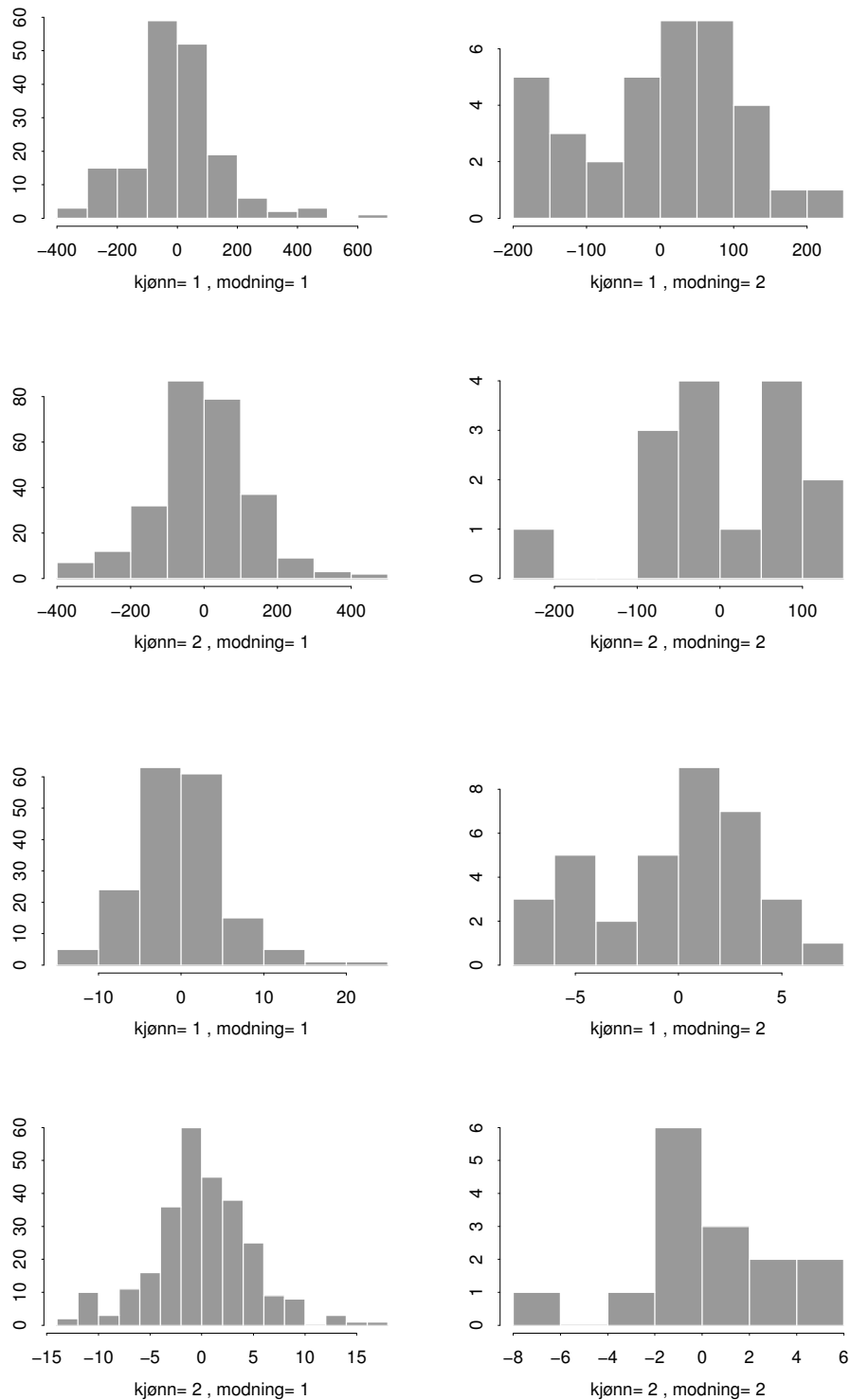
Figur D.2: Plott av modell for fixed effekt mot tid. Modell: tredjegradspolynom i samspill med modning nøstet innen kjønn. Fordelingen av data på hvert tidspunkt er vist med punkter og linjer som går gjennom midpunktene i histogrammer.



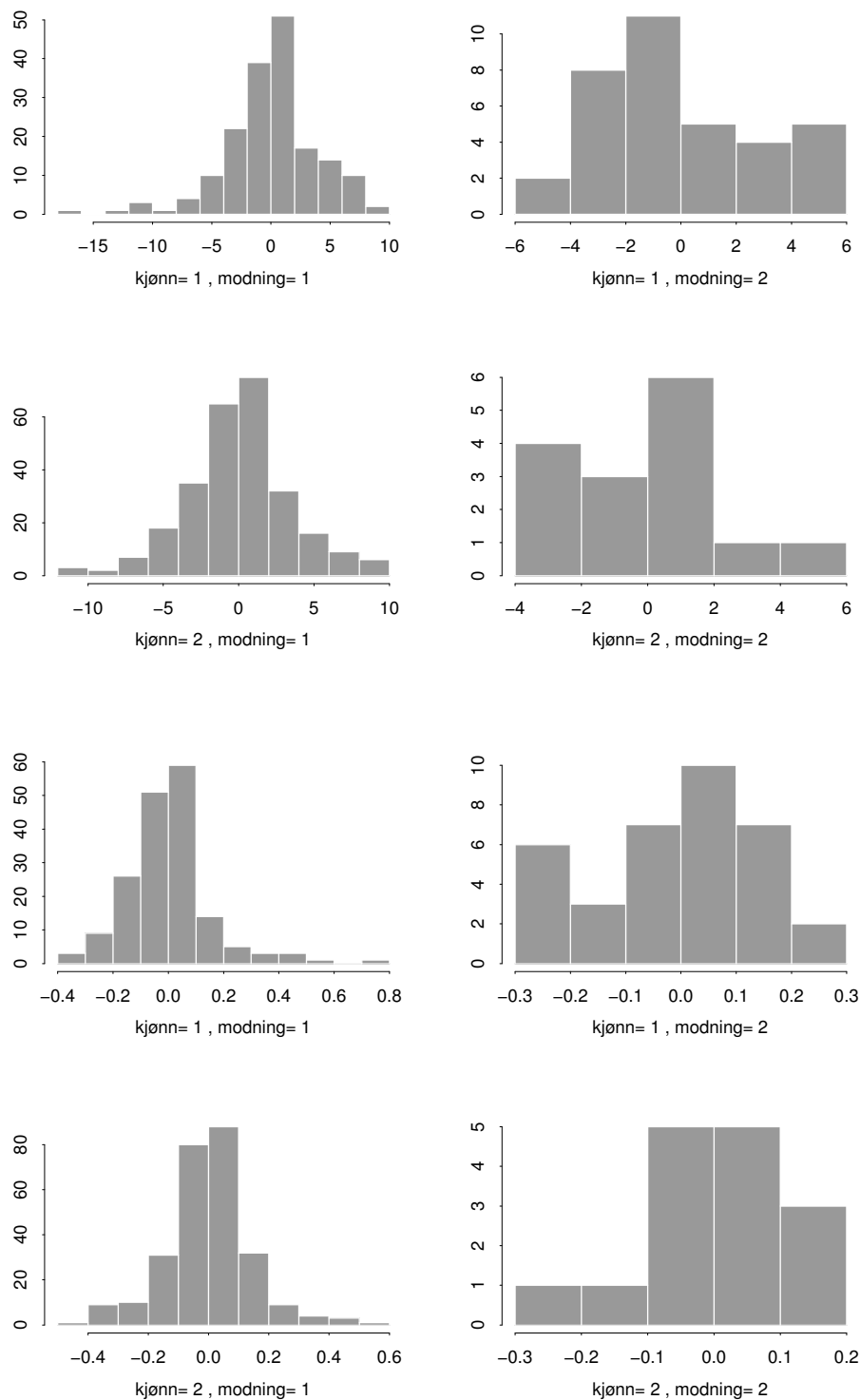
Figur D.3: Plott av residualer mot tid. Modell: tredjegradspolynom i samspill med modning nøstet innen kjønn (hanner=1, hunner=2, umoden=1 og moden=2).



Figur D.4: Individuelle plott (Individer valgt ved «random» funksjon). Modell: tredje-grads modell med modning nøstet innen kjønn (Vekt4-6). Komplette serier (lengde og vekt).



Figur D.5: Histogram av random effektene (1 og 2). Modell: Vektutviklingen forklart med tredjegradspolynom i samspill med modning nøstet innen kjønn.



Figur D.6: Histogram av random effektene (3 og 4). Modell: Vektutviklingen forklart med tredjegradspolynom i samspill med modning nøstet innen kjønn.

Register

- «powermodellen», 54
- 11-oxotestosterone, 24, 155
- alfanumerisk kode, 37
- anadrom, 16, 27
- ANCOVA, 7, 8, 63, 82, 93, 120
- ANOVA, 6, 7
- AR[1] prosess, 113, 118
- artefakt, 79
- autokorrelasjon, 83, 111
- autoregressiv stasjonær prosess, 118
- benzokain, 34
- Bertalaffy, 51
- beslutning, 24, 144
- beslutningsvinduet, 26
- bimodal, 65, 125, 133
- bimodal fordeling, 6, 19
- bløgging, 30, 39
- Bonferroni korleksjon, 5
- bootstrapping, 12, 86
- clustering, 41
- constraints, 27
- contr.poly, 121
- contr.treatment, 121
- costia, 34
- døgngrader, 17
- data-mining, 12, 79, 80
- determinant, 91
- devianser, 88
- diskriminantanalyse, 7, 8, 41, 83
- disordinal, 99
- dverghann, 23, 29, 36, 39
- eksponensiell vekst, 52
- EM-algoritmen, 12, 87
- evolusjonære stabile strategi, 27
- førstegangsgyter, 23
- fenotypisk plastisitet, 26, 27, 156
- feromonhypotesen, 23
- fettfinne, 16
- fitness, 25
- game teory, 27
- generalisert kvadratisk distanse, 38
- generasjonstiden, 25
- gls-rutinen, 82, 94
- Gompertz, 51
- gonade, 13
- Gonado Somatisk Indeks, 30, 36, 39
- gonadotropin, 24, 157
- growth models, 82
- GSI, 13, 30, 36, 39
- gyte-vandring, 24
- gyteklare, 21

- høstsmolt, 20
Hendersons metoder, 103
hierarki, 85, 136
histologi, 41
Hotelling-lawley, 102
- ikke-parametriske metoder, 5, 104
indikatorvariabler, 107
individeffekter, 52, 82, 105, 118
individuelle merker, 31, 33
informasjonskriterier, 88
iterativ, 57, 112
iteropar, 17
- kjønnsdrakt, 25
kjønnsmodning, 2, 4, 20, 24, 29, 157
kloridceller, 19
kompensasjonsvekst, 21
kondisjonsfaktor, 36, 38, 59, 120, 128, 133, 142
konfidens-ellipse, 38
korrelasjonsstruktur, 108
kovariansmatrise, 30, 31, 101, 108, 113–115
kovariat, 100, 150
- LD24:0, 4, 63, 129
likelihood-estimering, 85
likelihoodfunksjon, 109, 114
lineære blandede modeller, 82, 104
Linear Mixed-Effects, 104
livshistoriestrategi, 26
LME, 104, 143
lme-rutinen, 10, 82, 118
LmList-rutinen, 111
logistisk regresjon, 7, 8, 83
loglikelihood-forholds-test, 85, 88, 92
longitudinelle data, 7, 81, 111
- lysbehandling, 4, 29, 122
lysregime, 4, 57, 68, 155
lysstyring, 3
- MANCOVA, 8, 83, 100, 143, 150
MANOVA, 8, 97, 139, 143
manova-rutinen, 82
match, 13
mellomlaks, 23
merd, 21, 28
MLE, 92
MSW-laks, 29, 36, 39, 155
multinormalfordeling, 84, 90, 114, 147
multippel regresjon, 56, 93
- nøsting, 99, 143, 145, 189
Na-K-ATPase, 19
nlme-rutinen, 10, 118
Nonlinear Mixed-Effects, 105
ny måleteknologi, 14, 31, 83
- observasjonsstudier, 79, 156
oppdrettslaks, 28
ordinal, 99
- paneldata, 81
parr, 17, 32
Passiv Integrated Transponder, 31, 33
pellets, 28
Pillais spor, 102, 146
PIT-merker, 9
PIT-nummerleser, 37
plommeseckyngel, 17
polynomisk regresjon, 96
postsmoltmodne, 3, 23, 29, 36, 39, 60, 144
powerkurven, 52, 55
predator, 21
profilanalyse, 97

- protected ANOVA, 97
pseudoreplikasjoner, 82
- rammer, 9, 30
random effects, 105
reaksjonsnorm, 26, 27
REML, 92
repeterte målinger, 7, 81
Restricted maximum likelihood, 92
robust statistikk, 85
robuste metoder, 86
robusthet, 102
rognkorn, 39
Roys største rot, 102, 146
ryggdeformasjoner, 34
- sannsynlighetskvotetest, 87, 88
semelpar, 17
sentring, 94, 150
settefisk-anlegg, 6
SGR, 21, 36, 37, 51, 113, 154
simultansannsynlighetstettheten, 90
sjøalder, 23
smolt, 3, 134, 156
smoltalder, 23
smoltifisering, 17, 18
sneakfuckers, 23
snikere, 23
SNL, 4, 65, 125, 133
SNL+B, 4
sosial rangordning, 21
Specific Growth Rate, 36
spillteori, 27
stamfisk, 32
startfôringsfasen, 21
stasjonær autoregressiv prosess, 112
statistisk diagnostikk, 85
- storlaks, 23
- Taylorutvikling, 89
termisk vekstkoefisient, 53
territorial, 28
terskel, 156
tert, 23, 29, 36, 39, 144, 155
testobservator, 102, 146
tidsrekker, 7
tidsserie, 81
to-steg beslutningsmodell, 25
trade off, 26
- uteligger, 5, 7, 38
- våtvekt, 30, 39
variensfunksjon, 94, 122
variogram, 117
vekst modeller, 82
vekst-tabell, 51
vekstkurve, 10, 30, 51, 83, 112
villlaks, 28
- Wald-observator, 84, 89
Wilks lambda, 102, 146
- Zeitgeber, 20

Bibliografi

- Alm G. (1959), «Connection between maturity, size and age in fishes.», *Rep. Inst. Freshwat. Res. Drottingholm*, **vol. 40**, s. 5–145.
- Anderson M.J. (2001), «A new method for non-parametric multivariate analysis of variance», *Austral Ecology*, **vol. 26**, s. 32–46.
- Anderson-Sprecher R. (1994), «Model Comparisons and R^2 », *The American Statistician*, **vol. 48**, nr. 2, s. 113–117.
- Austreng E., Storebakken T., og Åsgård T. (1987), «Growth rate estimates for cultured atlantic salmon and rainbow trout», *Aquaculture*, **vol. 60**, s. 157–160.
- van Belle G. (2002), *Statistical Rules of Thumb*, JOHN WILEY & SONS, INC.
- Berg M. (1986), *Det norske lakse- og innlandsfiskets historie*, Universitetsforlaget.
- Blaalid G.E. (2005), «Sjømateksporten opp med to milliarder i 2004», *Norsk fiskeoppdrett*, **vol. 1**, s. 38–39.
- Brockwell P.J. og Davis R.A. (2002), *Introduction to Time Series and Forecasting*, Springer, 2. utg.
- Bureau D. (2000), «Pattern and cost of growth and nutrition deposition in fish and shrimp: Potential implications and applications», URL <http://www.uoguelph.ca/fishnutrition>.
- Cole D.A., Scott A. E. M. and Arvey, og Salas E. (1994), «How the power of MANOVA can both increase and decrease as a function of the Intercorrelations among the dependent variables», *Psychological Bulletin*, **vol. 115**, nr. 3, s. 465–474.
- Crawley M.J. (2002), *Statistical Computing: An Introduction to Analysis using S-plus*, JOHN WILEY & SONS, LTD.
- Diggle P.J. (1990), *Time Series: A Biostatistical Introduction*, Oxford Science Publications.

- Dobson A.J. (2002), *An introduction to generalized linear models*, Chapman & Hall/CRC.
- Endal H., Taranger G., Stefansson S.O., og Hansen T. (2000), «Effects of continuous additional light on growth and sexual maturity in Atlantic Salmon (*Salmo salar*), reared in sea cages», *Aquaculture*, **vol. 191**, s. 337–349.
- Fleming I.A. (1996), «Reproductive strategies of Atlantic salmon: ecology and evolution», *Reviews in Fish Biology and Fisheries*, **vol. 6**, s. 379–416.
- Forsberg O.I. (1995), *Farming of post-smolt Atlantic salmon (*Salmo salar* L.) in land-based flow-through tanks: studies of fish growth, metabolic rates, water quality and optimal production strategies*, doktorgradsavhandling, UiB.
- Garder M.L.G. (1976), «A review of factors which may influence the sea-age and maturation of Atlantic salmon *Salmo salar* L.», *J. Fish Biol.*, **vol. 9**, s. 289–327.
- Gjerde B. (1986), «Growth and Reproduction in Fish and Shellfish», *Aquaculture*, **vol. 57**, s. 37–55.
- Hair J. jr., Anderson R., Tatham R., og Black W. (1998), *Multivariate Data Analysis*, Prentice-Hall International, Inc.
- Hansen T. (red.) (1998), *Oppdrett av laksesmolt*, Landbruksforlaget.
- Hauck W.W.j. og Donner A. (1977), «Wald's Test Applied to Hypotheses in Logit Analysis», *Journal of the American Statistical Association*, **vol. 72**, nr. 360, s. 851–853.
- Heggberget T.G. (1988), «Timing of Spawning in Norwegian Atlantic Salmon (*Salmo salar*)», *Can. J. Fish Aquat. Sci.*, **vol. 45**, s. 845–849.
- Hoar W. (1988), *Fish Physiology*, vol. 11b, kap. 4, s. 275–325, Academic Press Inc.
- Hoening J.M. og Heisey D.M. (2001), «The Abuse of Power: The Pervasive Fallacy of Power Calculations for Data Analysis», *The American Statistician*, **vol. 55**, nr. 1, s. 19–24.
- Hunt S., Simpson T., og Wright R. (1982), «Seasonal changes in levels of 11-oxotestosterone and testosterone in the serum of male salmon, *Salmo salar* L., and their relationship to growth and maturation cycle», *J. Fish biol.*, **vol. 20**, s. 105–119.
- Hurlbert S.H. (1984), «Pseudoreplication and the design of ecological field experiments», *Ecological monographs*, **vol. 54**, nr. 2, s. 187–211.
- Høisæter T. (1989), «Kompendium i biologisk statistikk», *Teknisk rapport*, IMB. Universitetet i Bergen.

- Imslund A.K. (2001), *Growth mechanisms in fishes and the use of stochastic models to simulate growth and size variation*, Master-oppgave, Universitetet i Bergen.
- Ingebrigtsen O. (red.) (1982), *Akvakultur: Oppdrett av laksefisk*, NKS-Forlaget.
- Iversen A. og Kosmo J.P. (2004), «Kan vekstindeksene sammenlignes?», *Norsk Fiskeoppdrett*, **vol. 4**, s. 62–64.
- Jennions M.D. og Møller A.P. (2003), «A survey of the statistical power of research in behavioral ecology and animal behavior», *Behavioral Ecology*, **vol. 14**, nr. 3, s. 438–445.
- Jensen F. (1982), «Gudenålaksen», *Natur og Museum*, **vol. 3**, s. 1–24.
- Johnson D.H. (1999), «The Insignificance of Statistical Significance Testing», *Journal of Wildlife Management*, **vol. 63**, nr. 3, s. 763–772.
- Johnson R. og Wichern D. (2002), *Applied Multivariate Statistical Analysis*, Prentice Hall, 4. utg.
- Jones R. (1993), *Longitudinal Data with Serial Correlation: A State-space Approach*, Chapman & Hall.
- Kadri S., F. M.D., Metcalfe N.B., Huntingford F., og Thorpe J.E. (1996), «Differential patterns of feeding and resource accumulation in maturing and immature Atlantic salmon, *Salmo salar*», *Aquaculture*, **vol. 142**, s. 245–257.
- Kaufmann K.W. (1981), «Fitting and Using Growth Curves», *Oecologia*, **vol. 49**, s. 293–299.
- Klemetsen A., Amundsen P., Dempson J., Jonsson N. B. and Jonsson, O'Connell M., og Mortesen E. (2003), «Atlantic salmon (*Salmo salar* L.), brown trout (*Salmo trutta* L.) and Arctic charr (*Salvelinus alpinus* L.): a review of aspects of their life histories», *Ecology of freshwater fish*, **vol. 12**, s. 1–59.
- Kolding J. og Bergstad O.A. (1988), «Introduction to practical techniques of fisheries biology», Bergen.
- Kråkenes R., Hansen T., Stefanson S., og Taranger G.L. (1991), «Continuous light increase growth rate of Atlantic salmon (*Salmo salar* L.) postsmolts in sea cages», *Aquaculture*, **vol. 95**, s. 281–287.
- Kvalseth T.O. (1985), «Cautionary Note about R^2 », *The American Statistician*, **vol. 39**, nr. 4, s. 279–285.
- Lærum O.D. (1988), *Kroppsrhythmene og Lyset: En innføring i kronobiologi*, Alma Mater.

- Metcalf N.B. og Monaghan P. (2001), «Compensation for a bad start: grow now, pay later?», *TRENDS in Ecologi & Evolution*, **vol. 16**, nr. 5, s. 254–260.
- Nester M.R. (1996), «An Applied Statistician's Creed», *Applied Statistics*, **vol. 45**, nr. 4, s. 401–410.
- Nævdal G., Bjerk M. Ø. and Holm, Lerøy R., og Møller D. (1979), «Growth rate and age at first sexual maturity of Atlantic salmon smoltifying aged one and two years», *FiskDir. Skr. Ser. HavUnders.*, **vol. 17**, s. 11–17.
- Nævdal G., Holm M., Ingebrigtsen O., og Møller D. (1978a), «Variation in age at first spawning in Atlantic salmon», *Journal of the Fisheries Research Board of Canada*, **vol. 35**, nr. 1, s. 145–147.
- Nævdal G., Holm M., Lerøy R., og Møller D. (1978b), «Individual growth rate and age at first sexual maturity in Atlantic salmon», *FiskDir. Skr. Ser. HavUnders.*, **vol. 16**, s. 519–529.
- Olsen S.O. (1991), *Biochemical and physiological characterization of Atlantic salmon (Salmo salar) pituitary hormones*, Master-oppgave, University of Bergen.
- Oppedal F. (2002), *Influences of artificial light on Atlantic salmon (Salmon salar L.) in seawater*, doktorgradsavhandling, Universitetet i Bergen.
- Oppedal F., Juell J.E., Taranger G., og Hansen T. (2001), «Artificial light and season affects vertical distribution and swimming behavior of post-smolt Atlantic salmon in sea cages», *Journal of Fish Biologi*, **vol. 58**, s. 1570–1584.
- Pan J.X. og Fang K.T. (2002), *Growth Curve Models and Statistical Diagnostics*, Springer.
- Peterman R.M. (1990), «Statistical power analysis can improve fisheries reasearch and managment», *Can J.Fish.Aquat.Sci.*, **vol. 47**, s. 2–15.
- Pinheiro J. og Bates D. (2000), *Mixed-Effects Models in S and S-PLUS*, Springer.
- Ramsay J. og Silverman B. (1997), *Functional Data Analysis*, Springer.
- Roff D.A. (1992), *The Evolution of Life Histories*, Chapman & Hall.
- Rowe D., Thorpe J., og Shanks A. (1991), «Role of fat stores in the maturation of male Atlantic salmon (Salmo salar) parr», *Can. J. Fish. Aquat. Sci.*, **vol. 48**, s. 405– 413.
- Sandland R.L. og McGilchrist C.A. (1979), «Stochastic Growth Curve Analysis», *Biometrics*, **vol. 35**, s. 255–271.

- Saunders R.L., Duston J., og Benfey T.J. (1994), «Environmental and biological factors affecting growth dynamics in relation to smolting of Atlantic salmon (*Salmo salar* L.)», *Aquaculture and Fisheries Management*, **vol. 25**, s. 9–20.
- Scarnecchia D. (1983), «Age at Sexual Maturity in Icelandic Stocks of Atlantic Salmon (*Salmo salar*)», *Can. J. Fish. Aquat. Sci.*, **vol. 40**, s. 1456–1468.
- Schandy T. og Røed H. (red.) (2002), *Laksen: elvas konge*, Forlaget: Tom & Tom.
- Scheiner S.M. og Gurevitch J. (red.) (2001), *Design and Analysis of Ecological Experiments*, Oxford University Press.
- Searle S.R., Casella G., og McCulloch C.E. (1992), *Variance Components*, John Wiley & sons, inc.
- Skilbrei O.T. (1991), «Importance of Threshold Length and Photoperiod for the Development of Bimodal Length–Frequency Distribution in Atlantic Salmon (*Salmo salar*)», *Can. J. Fish. Aquat. Sci.*, **vol. 48**.
- Snedecor G. og Cochran W. (1989), *Statistical Methods*, Iowa State University Press/ AMES, 8. utg.
- Sokal R. og Rohlf (1981), *BIOMETRY*, W.H. Freeman and company.
- Stearns S.C. (1992), *The Evolution of Life Histories*, Oxford University Press.
- Stefansson S., Bjørnsson B.T., Hansen T., Haux C., Taranger G., og Sauders R. (1991), «Growth, parr-smolt transformation, and changes in growth hormone of Atlantic salmon reared under different photoperiods», *Can. J. Fish. Aquat. Sci.*, **vol. 48**, nr. 11, s. 2100–2108.
- Stefansson S.O. og Hansen T. (1989), «The effect of spectral composition on growth and smoltification in Atlantic salmon (*Salmo salar*) and subsequent growth in sea cages.», *Aquaculture*, **vol. 82**, s. 155–162.
- Sømme J.D. (1954), *Ørretboka*, Jacob Dywads Forlag.
- Taranger G.L. (1993), *Sexual maturation in Atlantic salmon, *Salmo salar* L.; aspects of environmental and hormone control*, doktorgradsavhandling, Universitetet i Bergen.
- Thorpe J.E. (1989), «Developmental variation in salmonid populations», *J. Fish Biol.*, **vol. 35**, s. 295–303, supplement A.
- (1994), «Reproductive strategies in Atlantic salmon, *Salmo salar* L.», *Aquaculture and Fisheries Management*, **vol. 25**, s. 77–87.

- Thorpe J.E., Talbot C., M.S. M., og Keay D. (1990), «Control of maturation in cultured Atlantic salmon, in pumped seawater tanks, by restricted food intake», *Aquaculture*, **vol. 86**.
- Troynikov V.S. (1998), «Probability density functions useful for parametrization of heterogeneity in growth and allometry data», *Bulletin of Mathematical biology*, **vol. 60**, s. 1099–1122.
- Venables W.N. og Ripley B. (2003), *Modern Applied Statistics with S*, Springer.
- Vestfossen K., Taranger G.L., Hansen T., og Stefansson S. (1991), «Growth and sexual maturation in reared Atlantic salmon(*Salmo salar*)», i Scott A., Sumpter J., Kime D., og Rolfe M. (red.), «Fourth International Symposium on the Reproductive Physiology of Fish», FishSymp 91, Sheffield.
- Wathne E. (1991), «Smolt Kompendiet», *Teknisk rapport*, Felleskjøpet.
- Wilcox R.R. (2001), *Fundamentals of Modern Statistical Methods: Substantially Improving Power And Accuracy*, Springer.
- Wooldridge J.M. (2003), *Introductory Econometrics : A Modern Approach*, Thomson South-Western, 2. utg.
- Wright P.S. (1992), «Adjusted P-Values for Simultaneous Inference», *Biometrics*, **vol. 48**, s. 1005–1013.
- Youngson A.F., McLay R. H.A. Wright, og Johnstone (1988), «Steroid Hormone Levels and Patters of Growth in the Early Part of the Reproductive Cycle of Adult Atlantic Salmon (*Salmo salar* L.)», *Aquaculture*, **vol. 69**, s. 145– 157.