



Ole Klungøy

Lineære modeller

- Av høyeste rang -

Notater

Forord

Jeg vil takke Anne Vedø og Arnfinn Schalm for inspirerende og gode diskusjoner gjennom hele prosessen det har vært å skrive dette notatet. De har med sin matematikkbakgrunn og innsikt vært av uvurderlig hjelp for å komme igjennom mange av bevisene for teoremene som de sentrale metodene bygger på. En god atmosfære oss i mellom har vært avgjørende for progresjonen og jeg synes vi har klart, også der forskjellige meninger har dukket opp, å bruke denne som konstruktiv drivkraft.

Innhold

1. Innledning	5
2. Bakgrunn	6
2.1 Modell.....	6
2.2 Estimering, generelt.....	6
2.3 Forvirring (entydighet).....	7
2.4 Mer forvirring (balanse).....	9
3. Viktige resultater fra regresjon (modell med full rang)	11
3.1 Generelt.....	11
3.2 Fordelingsegenskaper.....	13
3.3 Den generelle lineære hypotesen.....	16
4. Modeller som ikke har full rang	19
4.1 Generelt.....	20
4.2 Fordelingsegenskaper.....	23
4.3 Estimerbare funksjoner.....	26
4.4 Den generelle lineære hypotesen.....	28
5. Sammenligning av forskjellige modeller	31
5.1 Tilpasning av en konstant.....	31
5.2 1-veis klassifisering.....	32
5.3 2-veis kryssklassifisering, uten interaksjon.....	34
5.4 2-veis kryss klassifisering, med interaksjon.....	37
6. Oppklaring - hårfin balanse	43
6.1 Balanserte data.....	43
6.2 Eksempel.....	46
7. Mer oppklaring - lunefull entydighet	48
Generelt.....	49
Metode 1.....	50
Metode 2.....	50
Metode 3.....	51
Eksempel.....	52
Diskusjon.....	53
Punkt 1.....	53
Punkt 2.....	54
Punkt 3.....	54
Punkt 4.....	56
8. SAS GLM - kvadratsummer for enhver anledning	57
8.1 Forskjellige typer kvadratsummer i SAS GLM.....	58
8.2 Estimerbare funksjoner i SAS GLM.....	58
To eksempler.....	59
Hypoteser fremstilt ved de estimerbare funksjonene.....	60
Generelt.....	61
Eksemlene fortsatt.....	63
8.3 I beste mening.....	64
Type I.....	65
Type II.....	66
Type III.....	67
Type IV.....	69

6. Referanser	70
Apendiks A - Derivasjon m.h.p. vektorer	71
Apendiks B - Positivt definite matriser.....	72
Apendiks C - Matrisemultiplikasjon.....	73
De sist utgitte publikasjonene i serien Notater.....	74

1. Innledning

Høsten 2001 hadde noen av oss ved Seksjon 720 en kollokviegruppe med tema: Anvendelser av lineær algebra i lineære statistiske modeller. Forfatteren James E. Gentle sier i innledningen av læreboken sin fra 1998: "*Linear algebra is one of the most important mathematical and computational tools in the sciences*". Uten å ta stilling til en eventuell rankingliste, trykket vi likevel Gentle's utrop til vårt bryst (og i innledningen av notatet). Det er god reklame, men også en bekreftelse til de av oss som med jevne mellomrom føler at ett grunnkurs i lineær algebra, langt tilbake (med bare gode minner) er for lite. Det er likevel nok til å skjønne at lineær algebra er et virkelig slagkraftig verktøy for statistikk. Skremmende rader av summetegn blir enkle matriseprodukter og antall frihetsgrader blir rangen av en matrise. Når man først begynner på en slik ferd, kan den fort bli lang. Det er uendelig med elegante sammenhenger mellom matrisene, som fremstår med autoritet ved sine uthevede blokkbokstaver (som for å understreke sin viktighet). Så hvordan begrense omfanget?

"Most statistical analyses are based on linear models"; nok et beleilig sitat, dette fra forfatterne av en SAS manual for lineære modeller ([1]). Lineære statistiske modeller er altså det vi fokuserer på i dette notatet. Nærmere bestemt variansanalyse av ubalanserte data ved overparametriserte modeller. At dataene er ubalanserte betyr at det ikke er like mange observasjoner i de finest oppdelte kategoriene (cellene), ev. også at det er celler uten observasjoner. Overparametriserte modeller er modeller der antall parametre er større enn det antallet som entydig lar seg estimere fra de forskjellige cellene. Disse modellene har fått stor oppmerksomhet og selv om det er opplagte ulemper ved dem, har de fremdeles såpass stor utbredelse at alle statistikkpakker bruker dem. Det er innlysende at det er viktig å kunne analysere ubalanserte data. I praksis er jo data oftest ubalanserte, selv i planlagte forsøk med like mange observasjoner i hver kategori, kan frafall inntreffe underveis. Det er selvfølgelig mange bøker om dette emnet, flest med litt om ubalanserte data som ett av mange temaer, mens de vi har sett på er viet i sin helhet til det. Det er mange steder lineær algebra forenkler beregningene, men vi har funnet behov for å repetere (og lære mer) nødvendig teori som til dels er forutsatt kjent. Her er det også vanskelig å begrense mengden av interessant materiale, og i det hele tatt; hvorfor skrive om dette når det har vært gjort så bra, og så mange ganger før?

"... and most analyses of linear models can be performed by three SAS procedures: REG, ANOVA and GLM" er slutten av sitatet ovenfor, og herved det siste. Det er tatt med som slagord for programpakken som skal utføre selve analysen. Vi har valgt SAS siden dette er den vanligste å bruke i SSB, og vi konsentrerer oss om prosedyren GLM siden dette er den mest generelle for det vi ser på. For å håndtere overparametriserte modeller beregner SAS en generalisert invers. Dette, sammen med konsekvensene av at dataene er ubalanserte bidrar til mange muligheter for forvirring, allerede i de teoretiske resultatene. Det har vært stor uenighet i statistikk-miljøet om hva som er riktig måte å analysere slike data på, helt frem til 1980-tallet. Når det så kommer til algoritmene som skal løse dette i praksis, kan det være svært vanskelig å skjønne hva SAS gjør og hvorfor, spesielt hvis man bare har manualer til hjelp og ikke kjenner feltet fra før. Noe som opplagt gjør det vanskelig å lage gode manualbeskrivelser av det vi ser på, er at SAS trenger algoritmer som skal takle mange flere situasjoner enn våre, så metodene må være rasjonelle også i andre sammenhenger. Det kan være vanskelig å skille mellom de kravene modellene våre stiller og de egenskapene som er med for å ivareta generaliteten. Manualenes tekniske fremstilling om hvordan programmet skal brukes riktig, gir mindre plass til gode teoretiske begrunnelser for de forskjellige forutsetninger og valg programmet gjør. Dette er farlig i GLM sitt uoversiktelige landskap, og det er fort gjort å teste en helt annen hypotese enn den man tror man tester. Leser vi manualene feil? Røper vi bare vårt eget manual-kompleks som har fått utvikle seg ved stadig mislykkede forsøk på å finne svarene der?

Svarene på spørsmålene over forklarer innholdet av dette notatet. Inspirasjonen oppsto ved praktisk erfaring med SAS GLM og noen spørsmål som aldri fikk svar. Da var det naturlig og lete i litteraturen om lineære statistiske modeller og bl.a. i de referansene som står i manualene. Det igjen vekket behovet for repetisjon av lineær algebra og slik fortsatte det til et notat. Det er et subjektivt veiet gjennomsnitt av informasjon fra disse tre kildene. Spørsmålene begynte som uklarheter i manualene,

de blir belyst ved en oppsummerende gjennomgang av de lineære modellene (så konsentrert som mulig), og underveis når vi har behov for resultater fra lineær algebraen, presenteres disse, men uten bevis. Vi har ikke tatt med flere resultater enn akkurat de vi har trengt underveis (som på en naturlig måte begrenser omfanget), og de er presentert i den rekkefølgen behovene har meldt seg.

Som den rene ANOVA tabellen har vi ment at det er mulig å dekomponere forvirringen knyttet til denne typen analyse i forskjellige bidrag og forklare hvert bidrag for seg. "Forklaringsgraden til denne modellen" (dette notatet) er ikke ment å være 100 % og selvfølgelig avhengig av leserens frihetsgrader som tid til rådighet, men det er altså først og fremst ment som en oppsummering, en samling resultater og bakgrunn for bruk av SAS GLM. Det er skrevet først og fremst for den som har bruk for å kunne utføre variansanalyse på ubalanserte data, med lite forutsatte forkunnskaper. Det man ikke skjønner av resultater, kan man finne svar på i referansene. Det presenteres likevel mange formler og man bør ha noe kjennskap til modellene vi snakker om, og litt lineær algebra.

2. Bakgrunn

2.1 Modell

Med en lineær modell menes her en modell som kan gis på formen

$$(2.1) \quad \mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$$

der \mathbf{y} er en vektor med observasjoner, \mathbf{X} er en matrise der kolonnene er kovariater (regresjon), eller "dummy" variable med 1-ere og 0-ere som markerer hvilken gruppe observasjonen tilhører, \mathbf{b} er parametervektoren og \mathbf{e} er vektoren med feilledd. Vi skal nøye oss med å se på de tilfellene der \mathbf{b} er tenkt på som en vektor med ukjente konstanter ("fixed effects"), ikke tilfeldige variable. Det er enklest og har tydelig forbindelse med lineær algebraen. Modellen er lineær fordi parametervektoren inngår lineært i ligningen. For fullstendig spesifisering av modellen gjøres også antagelser om forventning og varians til feilleddene. Vi antar at forventningen til feilene er 0, at de har samme varians og er ukorrelerte, altså:

$$(2.2) \quad E(\mathbf{e}) = \mathbf{0}, \quad \text{var}(\mathbf{e}) = \sigma^2 \mathbf{I}$$

2.2 Estimering, generelt

Minste kvadraters estimering er en av flere estimeringsmetoder i denne modellen. Den kan lett generaliseres til tilfeller der feilene ikke har samme varians, er intuitiv og legger ingen forutsetninger på fordelingen til feilleddene. Den gir også samme parameterestimerer $\hat{\mathbf{b}}$ som f.eks. maximum likelihood estimering gir under forutsetning av normalfordelte feilledd.

For å finne parameterestimatene minimeres kvadratsummen av forskjellene mellom observasjonene og de forventede \mathbf{y} verdier, altså minimere flg. uttrykk:

$$\mathbf{e}'\mathbf{e} = [\mathbf{y} - E(\mathbf{y})]' [\mathbf{y} - E(\mathbf{y})] = (\mathbf{y} - \mathbf{X}\mathbf{b})' (\mathbf{y} - \mathbf{X}\mathbf{b}) = \mathbf{y}'\mathbf{y} - 2\mathbf{b}'\mathbf{X}'\mathbf{y} + \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b}$$

mhp parametervektoren \mathbf{b} . Ved å derivere mhp \mathbf{b} og sette lik 0 finnes minimum (for derivasjon mhp en vektor se appendix A). Ligningene som gir minimum kalles normalligningene og parametervektoren som tilfredsstill minimum er estimatet og kalles $\hat{\mathbf{b}}$. Normalligningene er gitt ved:

$$(2.3) \quad \mathbf{X}'\mathbf{X}\hat{\mathbf{b}} = \mathbf{X}'\mathbf{y}$$

Hvis \mathbf{X} har full rang har også $\mathbf{X}'\mathbf{X}$ full rang (se kapittel 3) og er invertibel (ikke-singulær). Da finnes den entydige løsningen for parameterestimaterne:

$$(2.4) \quad \hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$$

I de tilfellene der \mathbf{X} og $\mathbf{X}'\mathbf{X}$ ikke har full rang kan den ikke inverteres på vanlig måte, men en ikke entydig løsning kan likevel finnes ved å innføre en generalisert invers $\mathbf{X}'\mathbf{X}$ som betegnes med \mathbf{G} eller $(\mathbf{X}'\mathbf{X})^-$. En løsning til (2.3) er da gitt ved:

$$(2.5) \quad \mathbf{b}^0 = (\mathbf{X}'\mathbf{X})^- \mathbf{X}'\mathbf{y} = \mathbf{G}\mathbf{X}'\mathbf{y}$$

hvor symbolet \mathbf{b}^0 brukes for å indikere at dette ikke er et estimat, men en av mange mulige løsniner til normalligningene. Dette kommer vi tilbake til i kapittel 4.

2.3 Forvirring (entydighet)

Innføring av den generaliserte inversen \mathbf{G} har mange konsekvenser, også når det gjelder statistiske aspekter som estimering og testing av hypoteser i de lineære modellene. Hva kan estimeres og hvordan ser en relevant hypotesetest ut? Hva man skal gjøre med at det ikke nødvendigvis finnes en entydig løsning av normalligningene er en av de viktigste kildene til forvirring, selv med de nødvendige bøker og flere kilo med SAS manualer på skrivebordet. Estimerbarhet er nøkkelbegrepet her og er best illustrert ved et enkelt eksempel.

Eksempel 2.1:

Vi har flg. 1-veis klassifisering (enkleste formen av 2.1) med 3 grupper og 6 observasjoner fordelt som 3, 2, 1. Modellen kan skrives på formen:

$$(2.6) \quad y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

der $i = 1, 2, 3$ og $j = 1, \dots, n_i$. På vektorformen fra 2.1 kan den skrives som:

$$(2.7) \quad \mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$$

med

$$\mathbf{y} = \begin{pmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{31} \end{pmatrix} = \begin{pmatrix} 16 \\ 10 \\ 19 \\ 11 \\ 13 \\ 27 \end{pmatrix}, \quad \mathbf{e} = \begin{pmatrix} e_{11} \\ e_{12} \\ e_{13} \\ e_{21} \\ e_{22} \\ e_{31} \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

μ er tenkt på som et felles gjennomsnitt og $\alpha_1, \alpha_2, \alpha_3$ som de gruppespesifikke avvikene fra μ . Modellen (\mathbf{b}) har 4 parametere og bare 3 grupper å estimere dem fra. Vi ser at \mathbf{X} ikke har full

kolonnerang (de 3 siste kolonnene summeres til den første), så $X'X$ er ikke invertibel. Normalligningene (2.3) er gitt ved:

$$(2.8) \quad \begin{pmatrix} 6 & 3 & 2 & 1 \\ 3 & 3 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu^0 \\ \alpha_1^0 \\ \alpha_2^0 \\ \alpha_3^0 \end{pmatrix} = \begin{pmatrix} 96 \\ 45 \\ 24 \\ 27 \end{pmatrix}$$

som er ekvivalent med:

$$\begin{aligned} 6\mu^0 + 3\alpha_1^0 + 2\alpha_2^0 + \alpha_3^0 &= 96 \\ 3\mu^0 + 3\alpha_1^0 &= 45 \\ 2\mu^0 + 2\alpha_2^0 &= 24 \\ \mu^0 + \alpha_3^0 &= 27 \end{aligned}$$

Vi ser direkte at summen av de 3 nederste ligningene er lik den øverste ligningen. Dermed har vi uendelig mange løsninger av ligningene. 4 slike løsninger er vist i tabellen nedenfor:

PARAMETER	LØSNING			
	\mathbf{b}_1^0	\mathbf{b}_2^0	\mathbf{b}_3^0	\mathbf{b}_4^0
μ^0	16	14	27	-2982
α_1^0	-1	1	-12	2997
α_2^0	-4	-2	-15	2994
α_3^0	11	13	0	3009

Tabell 1: 4 løsninger av normalligningene

At en av disse løsningene fra tabellen ikke kan brukes som parameter estimat er innlysende, og grunnen er altså at vi har en overparametrisert modell, en modell med flere parametere enn ligninger som kan brukes for å estimere disse.

Ved å "fire" litt på kravene og i stedet for å kreve å estimere alle parametrene, nøye oss med å betrakte lineære funksjoner av flere parametre ser vi noe interessant. Tabellen nedenfor viser 2 lineære funksjoner av elementene i \mathbf{b}^0 for de forskjellige løsningene i tabell 1.

LINEÆR FUNKSJON	LØSNING FRA TABELL 1			
	\mathbf{b}_1^0	\mathbf{b}_2^0	\mathbf{b}_3^0	\mathbf{b}_4^0
$\frac{1}{2}(\alpha_2^0 + \alpha_3^0)$	3,5	5,5	-7,5	3 001,5
$\frac{1}{3}(\mu^0 + \alpha_1^0 + \alpha_2^0 + \alpha_3^0)$	22,33	26,33	0	2 006

Tabell 2: Noen lineære funksjoner av parametrene med forskjellig resultat for de forskjellige løsningene

Vi ser at disse 2 funksjonene gir forskjellige resultat for de forskjellige \mathbf{b}^0 fra tabell 1. Men for noen andre funksjoner får vi samme resultat for alle \mathbf{b}^0 , se tabellen nedenfor:

LINEÆR FUNKSJON	LØSNING FRA TABELL 1			
	\mathbf{b}_1^0	\mathbf{b}_2^0	\mathbf{b}_3^0	\mathbf{b}_4^0
$(\alpha_1^0 - \alpha_2^0)$	3	3	3	3
$(\mu^0 + \alpha_1^0)$	15	15	15	15
$\mu^0 + \frac{1}{2}(\alpha_2^0 + \alpha_3^0)$	19,5	19,5	19,5	19,5
$\frac{1}{2}(\alpha_2^0 + \alpha_3^0) - \alpha_1^0$	4,5	4,5	4,5	4,5

Tabell 3: Noen andre lineære funksjoner av parametrene med likt resultat for de forskjellige løsningene

Denne invariansegenskapen til disse funksjonene gjør at uansett hvilken løsning man bruker fås samme resultat som da kan brukes som estimat. F.eks. er den øverste funksjonen i tabell 3 en estimator på forskjellen i nivå mellom gruppe 1 og 2. Funksjoner med denne ønskede invariansegenskapen kalles **estimerbare funksjoner** og i kapittel 4 skal vi se hvordan man kan finne disse funksjonene og teste hypoteser om dem. •

2.4 Mer forvirring (balanse)

En annen hovedkilde til forvirring i disse lineære modellene, er skillet mellom tilfellene der man har like mange observasjoner i gruppene på det finest oppdelte nivået og der man ikke har det. Hvorfor man bør skille mellom disse, hvor forskjellene dukker opp i den statistiske inferensen, og det å betrakte det balanserte tilfellet (like mange observasjoner i gruppene) som et spesialtilfelle av det ubalanserte tilfellet (som er Searle sin angrepsvinkel i [3]), kommer vi tilbake til i kapittel 4, men bare for å vise denne forskjellen, er et eksempel nyttig.

Eksempel 2.2:

Vi tenker oss 6 observasjoner som er f.eks. produksjon fra planter av 3 typer, med 2 forskjellige gjødslingsmåter (behandling). Observasjonene kan fremstilles som i tabellen nedenfor og er et eksempel på balanserte data:

Plantetype	BEHANDLING (GJØDSLING)		Total
	1	2	
1	y_{11}	y_{12}	$y_{1.}$
2	y_{21}	y_{22}	$y_{2.}$
3	y_{31}	y_{32}	$y_{3.}$
Total	$y_{.1}$	$y_{.2}$	$y_{..}$

Tabell 4: Balanserte data

Her passer en 2-veis klassifisering, for enkelhets skyld uten interaksjon. Den modellen er gitt ved:

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$

hvor μ er det generelle gjennomsnitt for produksjon fra alle typer, mens α_i er effekten på produksjon av plantetype i , β_j er effekten på produksjon av behandling (gjødslingsmåte) j , og ε_{ij} er støyleddet (uavhengig, med 0 forventning og samme varians for alle). Til sammen gir dette 6 parametre (ser bort fra variansen i støyen som også er en ukjent parameter, men som ikke er med i parametervektoren). At modellen er overparametrisert kommer av at formen på den (additive effekter) gjør at de 6 gruppene man har å basere estimeringen på bare klarer å estimere 4 parametere entydig (som er lik rangen til \mathbf{X}) som er mindre enn antall ukjente parametre i modellen. Dette betyr altså at alle parametrene ikke kan estimeres, men som i forrige eksempel kan noen lineære funksjoner av parametrene estimeres. Vi antar nå at vi har samme antall observasjoner, men at vi i type 1 - behandling 1 har 2 observasjoner og ingen observasjon i type 3 - behandling 2. Da har vi ubalanserte data som kan fremstilles som nedenfor:

Plantetype	BEHANDLING (GJØDSLING)		Total
	1	2	
1	y_{111}, y_{112} (2)	y_{121} (1)	$y_{1.}$ (3)
2	y_{211} (1)	y_{221} (1)	$y_{2.}$ (2)
3	y_{311} (1)	- (0)	$y_{3.}$ (1)
Total	$y_{.1}$ (4)	$y_{.2}$ (2)	$y_{..}$ (6)

Tabell 5: Ubalanserte data (antall observasjoner i parantes)

Det viser seg at klassen estimerbare funksjoner er den samme i de to tilfellene. Det er svært lite som skiller de 2 tilfellene, men konsekvensene for estimeringen er stor. F.eks. er funksjonen $\beta_1 - \beta_2$ estimerbar og tabellen nedenfor viser estimatoren for denne funksjonen (estimeringsmetoden er BLUE - best linear unbiased estimator) i balansert tilfelle og i ubalansert tilfelle:

ESTIMERBAR FUNKSJON	ESTIMATOR	
	Balanserte data (Tabell 4)	Ubalanserte data (Tabell 5)
$\beta_1 - \beta_2$	$\bar{y}_{.1} - \bar{y}_{.2}$	$-\frac{1}{7}(4y_{1.} + 3y_{2.} + 6y_{3.} - 6y_{.1})$

Tabell 6: Estimator for balansert og ubalansert tilfelle

Vi ser stor forskjell på estimatoren i de to tilfellene. Estimatoren i det balanserte tilfellet er intuitiv, det er rimelig å estimere nivåforskjell i kolonneeffekten (behandling), altså forskjell på de to behandlingene, med differansen av kolonnegjennomsnittene. Derimot er det ikke lett å tolke resultatet i det ubalanserte tilfellet. •

Dette eksempelet viser at om dataene er balanserte eller ikke kan ha stor betydning for den statistiske analysen.

3. Viktige resultater fra regresjon (modell med full rang)

Først betrakter vi modellen med full rang. Da slipper man unna forvirringen som skyldes overparametriseringen mens mange av de sentrale begrepene kan innføres. Dette er ikke ment som noen oversikt over regresjonsanalysen, (resultater fra vanlig lineær regresjon forutsettes kjent) bare et utvalg av resultater som er basert på lineær algebra og som ligner på de resultatene som gjelder for den overparametriserte modellen, med vekt på hypotesetesting.

3.1 Generelt

Den vanlige regresjonsmodellen:

$$(3.1) \quad y_i = b_0 x_{i0} + \dots + b_k x_{ik} + \varepsilon_i$$

passer inn i (2.1) med

$$(3.2) \quad \mathbf{X} = \begin{pmatrix} x_{10} & x_{11} & \dots & x_{1k} \\ x_{20} & x_{21} & & x_{2k} \\ \vdots & \vdots & & \vdots \\ x_{N0} & x_{N1} & \dots & x_{Nk} \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} b_0 \\ \vdots \\ b_k \end{pmatrix}$$

der altså kolonnene i \mathbf{X} er kovariatene. Nå har altså $\mathbf{X}'\mathbf{X}$ full rang (nesten alltid), er dermed invertibel og parameterestimatet er gitt ved ligning (2.4). En viktig egenskap med denne estimatoren er at med variansbetingelsen (2.2) oppfylt er dette estimatet "best, linear, unbiased, estimator" (b.l.u.e.), se kapittel 4. Det er lineært i observasjonene, og best i den forstand at blant de forventningsrette estimatorene har den minst varians. Forventningsrettheten og uttrykket for variansen følger direkte:

$$(3.3) \quad \begin{aligned} E(\hat{\mathbf{b}}) &= E\{(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}\} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{b}, \\ \text{var}(\hat{\mathbf{b}}) &= E\left\{(\hat{\mathbf{b}} - E\hat{\mathbf{b}})(\hat{\mathbf{b}} - E\hat{\mathbf{b}})'\right\} = E\left\{(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'(\mathbf{y} - E\mathbf{y})(\mathbf{y}' - E\mathbf{y}')\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\right\} \\ &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'E(\mathbf{e}\mathbf{e}')\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = (\mathbf{X}'\mathbf{X})^{-1} \sigma^2 \end{aligned}$$

Som vanlig brukes notasjonen $\hat{\mathbf{y}} = \widehat{E(\mathbf{y})} = \mathbf{X}\hat{\mathbf{b}}$ for estimatet av forventningen til \mathbf{y} (predikerte verdier). Da blir avvikene mellom de observerte y_i -ene og de predikerte verdiene \hat{y}_i gitt ved

$$(3.4) \quad \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\hat{\mathbf{b}} = \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} = [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}']\mathbf{y}$$

og kvadratsummen av avvikene, heretter kalt residualfeilkvadratsummen, med symbol SSE :

$$(3.5) \quad \begin{aligned} SSE &= \sum_{i=1}^N (y_i - \hat{y}_i)^2 = (\mathbf{y} - \hat{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}}) = \mathbf{y}'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}']\mathbf{y} \\ &= \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} = \mathbf{y}'\mathbf{y} - \hat{\mathbf{b}}'\mathbf{X}'\mathbf{y} \end{aligned}$$

siden $[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']$ er symmetrisk og idempotent (blir uforandret om ganget med seg selv). For å estimere σ^2 tas utgangspunkt i SSE . Da trengs først et fordelingsresultat om kvadratiske former ([3], s 55):

Teorem 3.1: Når $\mathbf{x} \sim (\boldsymbol{\mu}, \mathbf{V})$ er $E(\mathbf{x}'\mathbf{A}\mathbf{x}) = \text{tr}(\mathbf{A}\mathbf{V}) + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}$

Dette brukes til å beregne forventningen til SSE . Altså med $\mathbf{y} \sim (\mathbf{X}\boldsymbol{\beta}, \mathbf{I}\sigma^2)$ får vi flg. forventning når vi bruker Teorem 3.1;

$$\begin{aligned}
 E(SSE) &= \text{tr}[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{I}\sigma^2 + \mathbf{b}'\mathbf{X}'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{X}\mathbf{b} \\
 (3.6) \quad &= r[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\sigma^2 \\
 &= [N - r(\mathbf{X})]\sigma^2
 \end{aligned}$$

Det siste leddet i første linje faller bort siden $[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{X} = \mathbf{0}$. Videre brukes at trace til en idempotent matrise er lik dets rang ([4], s 320). Deretter kan flg. lemma brukes ([4], s 198):

Lemma 3.1: Når \mathbf{M} idempotent av dimensjon n gjelder flg.: $r(\mathbf{I} - \mathbf{M}) = n - r(\mathbf{M})$

siden $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ er idempotent og av dimensjon $N \times N$. Det som gjenstår er å vise at $r(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') = r(\mathbf{X})$. Da trenger vi et generelt resultat om rangen til produktet av 2 matriser \mathbf{A} og \mathbf{B} ([4], s 196):

Teorem 3.2: $r_{\mathbf{AB}} \leq \min(r_{\mathbf{A}}, r_{\mathbf{B}})$

Alle omformingene vi trenger her baseres på at \mathbf{X} har full kolonnerang, (husk at kolonnerangen som er antall lineært uavhengige kolonner er lik radrang og full kolonnerang betyr at alle kolonnene er lineært uavhengige). Her er altså $r(\mathbf{X}) = k + 1$, og $\mathbf{X}'\mathbf{X}$ er invertibel. Ved å bruke at

$$\mathbf{X} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} \text{ får vi ved Teorem 3.2 at: } r(\mathbf{X}) = r(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}) \leq r(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') \leq r(\mathbf{X})$$

som viser at $r(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') = r(\mathbf{X})$. Dette resultatet kan også brukes for å vise at

$$r(\mathbf{X}'\mathbf{X}) = r(\mathbf{X}\mathbf{X}') = r(\mathbf{X}). \text{ Dette kommer igjen flere steder. Teorem 3.2 brukt to ganger gir både at } r(\mathbf{X}'\mathbf{X}) \leq r(\mathbf{X}) \text{ og ved samme omskriving av } \mathbf{X} \text{ som ovenfor at } r(\mathbf{X}) = r(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}) \leq r(\mathbf{X}'\mathbf{X})$$

slik at $r(\mathbf{X}'\mathbf{X}) = r(\mathbf{X})$ er vist. Helt tilsvarende har vi at $r(\mathbf{X}\mathbf{X}') \leq r(\mathbf{X})$ og ved omskrivingen

$$\mathbf{X} = (\mathbf{X}\mathbf{X}')\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \text{ gjelder at } r(\mathbf{X}) \leq r(\mathbf{X}\mathbf{X}') \text{ og dermed har vi vist at } r(\mathbf{X}\mathbf{X}') = r(\mathbf{X}).$$

Fra (3.6) har vi nå fått et forventningsrett estimat for σ^2 ved:

$$(3.7) \quad \hat{\sigma}^2 = \frac{SSE}{N - r(\mathbf{X})} = \frac{SSE}{N - k - 1}$$

For å kunne splitte opp de forskjellige kildene til variasjon i dataene innføres den totale kvadratsummen av observasjonene, forkortet SST , og gitt ved:

$$(3.8) \quad SST = \mathbf{y}'\mathbf{y} = \sum_{i=1}^N y_i^2$$

Forskjellen mellom den totale kvadratsummen SST og SSE kalles SSR og er altså den delen av den totale kvadratsummen som modellen bidrar med (regresjonskvadratsummen), slik at vi har:

$$(3.9) \quad SSR = SST - SSE = \hat{\mathbf{b}}'\mathbf{X}'\mathbf{y}$$

med SSE fra (3.5). Denne oppsplittingen er grunnlaget for den vanlige ANOVA-tabellen, bortsett fra at da brukes ofte formen som er korrigert for gjennomsnittet hvor fra hvert element i \mathbf{X} matrisen kolonnegjennomsnittet er trukket fra. Disse 2 formene er i 1-1 forbindelse med hverandre, men sistnevnte er fordelaktig når man vil sammenligne modellene med og uten intercept.

Modellen uten kovariater, $y_i = b_0 + \varepsilon_i$, altså å tilpasse dataene til en konstant, har en spesiell rolle og har fått eget navn på kvadratsumbidraget, SSM . Vi ser lett at siden $\mathbf{X} = \mathbf{1}$ (en kolonne av bare 1-ere) blir $\hat{b}_0 = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} = \bar{y}$. Da blir SSR fra (3.9): $SSR = \hat{b}_0 \mathbf{X}'\mathbf{y} = \bar{y} \mathbf{1}'\mathbf{y} = N\bar{y}^2$ som er kvadratsumbidraget fra den enkleste mulige modellen og kalles SSM , og er "korreksjon for gjennomsnittet":

$$(3.10) \quad \begin{aligned} SSM &= N\bar{y}^2 \\ SSR_m &= SSR - SSM \\ SST_m &= SST - SSM \end{aligned}$$

SSR_m og SST_m er altså regresjonskvadratsummen og totalkvadratsummen korrigert for gjennomsnittet.

3.2 Fordelingsegenskaper

Nå innfører vi antagelsen om normalfordelt støy. Det er forutsetningen for at de forskjellige kvadratsummene (SSR , SSE , SSM , SSR_m , SST) er χ^2 fordelt og at F fordelingen kan brukes i testing av diverse hypoteser. Vi antar altså:

$$(3.11) \quad \mathbf{e} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$$

Da blir fordelingen til \mathbf{y} :

$$(3.12) \quad \mathbf{y} \sim N(\mathbf{X}\mathbf{b}, \sigma^2 \mathbf{I})$$

Siden $\hat{\mathbf{b}}$ er en lineær funksjon av \mathbf{y} er denne også normalfordelt:

$$(3.13) \quad \hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \sim N(\mathbf{b}, (\mathbf{X}'\mathbf{X})^{-1}\sigma^2)$$

fra (3.3). Videre gjelder:

$$(3.14) \quad \hat{\mathbf{b}} \text{ og } \hat{\sigma}^2 \text{ er uavhengige}$$

For å se dette brukes flg. teorem ([3], s 59):

Teorem 3.3: Når $\mathbf{x} \sim N(\boldsymbol{\mu}, \mathbf{V})$ er $\mathbf{x}'\mathbf{A}\mathbf{x}$ og $\mathbf{B}\mathbf{x}$ uavhengige hvis og bare hvis $\mathbf{BVA} = \mathbf{0}$

Siden $SSE = \mathbf{y}'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y}$ fra (3.5), som er en kvadratisk form i \mathbf{y} , og \mathbf{y} er normalfordelt (fra (3.12)) vil teoremet si at SSE og $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \hat{\mathbf{b}}$ er uavhengig hvis og bare hvis $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] = \mathbf{0}$ noe vi ser den er, altså gjelder (3.14).

For å se at SSE er χ^2 trenger vi nok et teorem ([4], s 57):

Teorem 3.4: Når $\mathbf{x} \sim N(\boldsymbol{\mu}, \mathbf{V})$ er $\mathbf{x}'\mathbf{A}\mathbf{x} \sim \chi^2[r(\mathbf{A}), \frac{1}{2}\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}]$ hvis og bare hvis \mathbf{AV} er idempotent

der apostrofen i $\chi^2[.,.]$ markerer at fordelingen er ikke-sentral og første parameter i parantesen er antall frihetsgrader og andre parameter er ikke-sentralitetsparameter. Siden $SSE = \mathbf{y}'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y} = \mathbf{y}'\mathbf{P}\mathbf{y}$ er kvadratisk i \mathbf{y} med \mathbf{P} som idempotent og $\text{var}(\mathbf{y}) = \sigma^2\mathbf{I}$ er også $(\frac{1}{\sigma^2})\mathbf{P}\sigma^2\mathbf{I}$ idempotent og dermed gjelder teoremet med flg. fordeling på SSE/σ^2 :

$$(3.15) \quad SSE/\sigma^2 \sim \chi^2[r(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'), \mathbf{b}'\mathbf{X}'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{X}\mathbf{b}/2\sigma^2] = \chi^2_{N-r}$$

der $r = r(\mathbf{X})$. Forenklingen her er pga. $[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{X} = \mathbf{0}$ så fordelingen blir sentral i dette tilfellet. Dessuten så vi i (3.6) at $r[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] = N - r(\mathbf{X})$.

For å beskrive SSR , dens fordeling og uavhengigheten mellom SSR og SSE trenger vi et siste teorem ([4], s 59):

Teorem 3.5: Når $\mathbf{x} \sim N(\boldsymbol{\mu}, \mathbf{V})$, er $\mathbf{x}'\mathbf{A}\mathbf{x}$ og $\mathbf{x}'\mathbf{B}\mathbf{x}$ uavhengige hvis og bare hvis $\mathbf{AVB} = \mathbf{0}$

Fra (3.9) har vi at $SSR = \hat{\mathbf{b}}'\mathbf{X}'\mathbf{y} = \mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ som også er en kvadratisk form i \mathbf{y} . Siden $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ er idempotent og ganget med $\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ gir $\mathbf{0}$ er, ved Teorem 3.5 SSR uavhengig av SSE og ved Teorem 3.4 har den flg. fordeling:

$$(3.16) \quad \begin{aligned} SSR/\sigma^2 &\sim \chi^2 \left[r(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'), \mathbf{b}'\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\mathbf{b}/2\sigma^2 \right] \\ &= \chi^2 \left[r, \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b}/2\sigma^2 \right] \end{aligned}$$

altså en ikke-sentral χ^2 fordeling, med $r = r(\mathbf{X})$ frihetsgrader, og ikke-sentralitetsparameter andre parameter i parantesen.

Tilsvarende har vi for SSM , at $SSM = N\bar{y}^2 = \mathbf{y}'N^{-1}\mathbf{1}\mathbf{1}'\mathbf{y}$, altså en kvadratisk form i \mathbf{y} , den også.

$N^{-1}\mathbf{1}\mathbf{1}'$ er idempotent, ganget med $\sigma^2\mathbf{I}$ er den fremdeles idempotent og det gir oss fordelingen ved Teorem 3.4. At den ganget med $\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ gir $\mathbf{0}$, som ved Teorem 3.5 gir uavhengigheten av SSE , krever noen argumenter for å se. Vi bruker først at alle modeller som assosieres med SSM (der SSM er av interesse) har intercept. Da er første kolonne i \mathbf{X} en kolonne med 1-ere. Hvis vi tenker oss \mathbf{X} med dimensjon $N \times (k+1)$ som f.eks.

$$\mathbf{X} = (\mathbf{1} \quad \mathbf{x}_1 \quad \cdots \quad \mathbf{x}_k) \text{ og dermed } \mathbf{X}' = \begin{pmatrix} \mathbf{1}' \\ \mathbf{x}_1' \\ \vdots \\ \mathbf{x}_k' \end{pmatrix}$$

bruker vi at $\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{X}'$ og at et matriseprodukt kan betraktes som en ny matrise der radene er hver rad i venstrematrisen ganget med hele matrisen til høyre (og tilsvarende gjelder for kolonnene, at kolonnene i resultatmatrisen er hele venstrematrisen ganget med hver kolonne i høyrematrisen). Da kan vi se på uttrykket for \mathbf{X}' som:

$$\begin{pmatrix} \mathbf{1}' \\ \mathbf{x}_1' \\ \vdots \\ \mathbf{x}_k' \end{pmatrix} = \mathbf{X}' = \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \begin{pmatrix} \mathbf{1}' \\ \mathbf{x}_1' \\ \vdots \\ \mathbf{x}_k' \end{pmatrix} \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \begin{pmatrix} \mathbf{1}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\ \mathbf{x}_1'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\ \vdots \\ \mathbf{x}_k'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \end{pmatrix}$$

og hvis dette skal stemme må det stemme for alle radene og dermed har vi at $\mathbf{1}' = \mathbf{1}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Dette resultatet brukes for å vise uavhengigheten mellom SSM og SSE . Betingelsen fra Teorem 3.5 er oppfylt siden $N^{-1}\mathbf{1}\mathbf{1}'(\sigma^2\mathbf{I})(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') = (N/\sigma^2)^{-1}(\mathbf{1}\mathbf{1}' - \mathbf{1}\mathbf{1}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') = \mathbf{0}$. Fordelingen er fra Teorem 3.4:

$$(3.17) \quad \begin{aligned} SSM/\sigma^2 &\sim \chi^2 \left[r(N^{-1}\mathbf{1}\mathbf{1}'), \mathbf{b}'\mathbf{X}'N^{-1}\mathbf{1}\mathbf{1}'\mathbf{X}\mathbf{b}/2\sigma^2 \right] \\ &= \chi^2 \left[1, (\mathbf{1}'\mathbf{X}\mathbf{b})^2/2N\sigma^2 \right] \end{aligned}$$

SSR_m og SST er også χ^2 fordelt på samme måte.

Ved å bruke definisjonen av en ikke-sentral F -fordeling ([3], s 51), brøken av en ikke-sentral χ^2 fordeling delt på sine frihetsgrader i teller og en sentral χ^2 fordeling delt på sine frihetsgrader i nevner kan vi nå uttrykke to relevante F -observatorer for testing av hypoteser ved:

$$(3.18) \quad F(R) = \frac{SSR/r}{SSE/(N-r)} \sim F(r, N-r, \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b} / 2\sigma^2)$$

$$F(M) = \frac{SSM/1}{SSE/(N-r)} \sim F(1, N-r, (\mathbf{1}'\mathbf{X}\mathbf{b})^2 / 2N\sigma^2)$$

ved å bruke (3.15), (3.16) og (3.17). Under visse null hypoteser er disse F -fordelingene sentrale og dermed egnet for testing av disse hypotesene.

3.3 Den generelle lineære hypotesen

Fire vanlige hypoteser som er interessante i regresjonsmodellen vår kan beskrives som (følger [3] sin fremstilling, s 110): (i) $H: \mathbf{b} = \mathbf{0}$, altså hypotesen om at alle parametrene er 0 (forkastning betyr at minst en parameter er signifikant forskjellig fra 0), (ii) $H: \mathbf{b} = \mathbf{b}_0$, som er hypotesen om at alle parametrene har en bestemt verdi $b_i = b_{i0}, i = 0, 1, \dots, k$, (iii) $H: \boldsymbol{\lambda}'\mathbf{b} = m$, hypotesen om at en lineærkombinasjon av parametrene er lik en bestemt verdi, og sist (iv) $H: \mathbf{b}_q = \mathbf{0}$, at en undermengde av parametrene, q stykker, $q < k$ er 0. Alle disse hypotesene (og mange flere) lar seg uttrykke på den generelle formen:

$$(3.19) \quad H: \mathbf{K}'\mathbf{b} = \mathbf{m}$$

der \mathbf{b} er den $k+1$ dimensjonale parametervektoren, \mathbf{K}' er en vilkårlig matrise av dimensjon $s \times (k+1)$ (uttrykkes som en transponert matrise, da stemmer notasjonen også for $s=1$ med vanlig notasjon for radvektor) og \mathbf{m} er en s dimensjonal vektor med spesifiserte konstanter. Eneste kravet til \mathbf{K}' er at den skal ha full rad-rang, altså at $r(\mathbf{K}') = s$. Dette betyr ikke noe annet enn at de lineære funksjonene av \mathbf{b} som utgjør hypotesene skal være lineært uavhengige. Det er et rimelig krav siden det som et eksempel betyr at hvis den ene hypotesen kan skrives $b_1 - b_2 = 2$ er det ikke noe vits i å ha med hypotesen $2b_1 - 2b_2 = 4$. Intuitivt bør også hypotesene som kan betraktes som et sett med ligninger ikke være inkonsistente, altså at man ikke har med hypotesen $b_1 - b_2 = 2$ og samtidig en annen hypotese $b_1 - b_2 = 3$. Som nyttig repetisjon er definisjonen av konsistens i ligningssett tatt med ([4], s 229):

Definisjon 3.1: De lineære ligningene $\mathbf{A}\mathbf{x} = \mathbf{y}$ er konsistente hvis de lineære sammenhengene mellom radene i \mathbf{A} også gjelder mellom de korresponderende elementene i \mathbf{y} .

Det viktige med denne definisjonen er at konsistente ligninger lar seg løse, mens inkonsistente ligninger ikke lar seg løse. Vi kommer tilbake til den flere ganger senere. Det er viktig å legge merke til at den ikke krever at det skal være noen lineære sammenhenger mellom radene i \mathbf{A} , men hvis det er, må de samme sammenhengene også gjelde for elementene i \mathbf{y} . Hvis \mathbf{A} har full radrang så finnes ikke lineære sammenhenger mellom radene og da er ligningene konsistente for alle \mathbf{y} .

Andvendt på (3.19) er poenget at de lineære hypotesene ikke skal inneholde samme informasjon (innbyrdes) og ikke være selvmotsigende. Men når kravet at \mathbf{K}' har full radrang er oppfylt er begge disse to egenskapene oppfylt. Hver hypotese kan ikke uttrykkes som en lineærkombinasjon av de andre, og konsistensen er oppfylt for alle \mathbf{m} . Den generelle F -observatoren for å teste (3.19) lar seg nå uttrykke. Fra (3.11) og (3.12) har vi at

$$(3.20) \quad \mathbf{K}'\hat{\mathbf{b}} - \mathbf{m} \sim N[\mathbf{K}'\mathbf{b} - \mathbf{m}, \mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}\sigma^2]$$

Ved å lage en kvadratisk form i $\mathbf{K}'\hat{\mathbf{b}} - \mathbf{m}$, og med $[\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}]^{-1}$ som matrisen, vil denne kvadratiske formen ha en χ^2 fordeling. Hvis vi kaller denne kvadratiske formen for Q , slik at:

$$(3.21) \quad Q = (\mathbf{K}'\hat{\mathbf{b}} - \mathbf{m})' [\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}]^{-1} (\mathbf{K}'\hat{\mathbf{b}} - \mathbf{m})$$

vil, ved Teorem 3.4 vi ha flg. fordeling på $\frac{Q}{\sigma^2}$:

$$(3.22) \quad \frac{Q}{\sigma^2} \sim \chi^2 \left[s, (\mathbf{K}'\hat{\mathbf{b}} - \mathbf{m})' [\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}]^{-1} (\mathbf{K}'\hat{\mathbf{b}} - \mathbf{m}) / 2\sigma^2 \right]$$

For det første er det viktig å merke seg at matrisen i Q alltid eksisterer, altså at $[\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}]$ er ikke-singulær. For å se dette gjør vi bruk av begrepet positivt definit (se Apendiks B), og flg. resultater ([3], s 35, 36 og [4] s 197):

Lemma 3.2: En symmetrisk matrise \mathbf{A} er positivt definit hvis og bare hvis alle dens "principal leading minors" har positive determinanter.

Determinanten til en matrise kan uttrykkes ved lavereordens determinanter ("minors") og "principal leading minors" er disse determinantene med tilhørende matrise som har diagonalen sin langs diagonalen i \mathbf{A} . Dette lemmaet gir flg. resultat:

Korolar: Positivt definite matriser er ikke-singulære (ikke nødvendigvis motsatt)

Dette er fordi at determinanten til hele \mathbf{A} er en av \mathbf{A} 's mange "principal leading minors" og hvis denne er positiv er \mathbf{A} nødvendigvis ikke-singulær (invertibel). Legg også merke til at selv om determinanten til \mathbf{A} er positiv trenger ikke nødvendigvis alle de lavere ordens determinantene være positive og derfor er ikke nødvendigvis alle ikke-singulære matriser positivt definite. Med dette korolaret er det nok å vise at $[\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}]$ er positivt definit. Da trengs tre lemmaer til:

Lemma 3.3: $\mathbf{A}\mathbf{A}'$ er positivt definit når \mathbf{A} har full rad-rang. Tilsvarende er $\mathbf{A}'\mathbf{A}$ positivt definit når \mathbf{A} har full kolonnerang.

Dette lemmaet brukes til å se at $\mathbf{X}'\mathbf{X}$ er positivt definit (\mathbf{X} har full kolonne-rang). Videre trengs:

Lemma 3.4: En symmetrisk matrise \mathbf{A} er positivt definit hvis og bare hvis den kan skrives som $\mathbf{P}'\mathbf{P}$ for en ikke-singulær \mathbf{P} .

Siden $\mathbf{X}'\mathbf{X}$ er symmetrisk og positivt definit finnes altså ved Lemma 3.4 en ikke-singulær \mathbf{P} slik at $\mathbf{X}'\mathbf{X} = \mathbf{P}'\mathbf{P}$. Da gjelder: $\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K} = \mathbf{K}'(\mathbf{P}'\mathbf{P})^{-1}\mathbf{K} = \mathbf{K}'\mathbf{P}^{-1}(\mathbf{P}')^{-1}\mathbf{K} = \mathbf{K}'\mathbf{P}^{-1}(\mathbf{P}^{-1})'\mathbf{K}$ (regel for inverse av produktet av to matriser og invers av transponert matrise). Siste lemmaet vi trenger er:

Lemma 3.5: Multiplikasjon med en ikke-singulær matrise forandrer ikke rangen.

Siden vi vet at \mathbf{K}' har full rad-rang ($r(\mathbf{K}') = s$) så sier Lemma 3.5 at rangen ikke endres ved å gange med \mathbf{P}^{-1} . Dermed har vi at $\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K} = \mathbf{K}'\mathbf{P}^{-1}(\mathbf{P}^{-1})'\mathbf{K} = \mathbf{B}\mathbf{B}'$ for en \mathbf{B} med full rad-rang (s), og ved Lemma 3.3 er den da positivt definit og dermed ikke-singulær.

Nå kan vi utlede F -observatoren for den generelle hypotesen. Da må vi først vise uavhengighet mellom Q fra (3.21) og SSE vha Teorem 3.5. Det gjøres ved å uttrykke begge som kvadratiske former i samme normalfordelte variabel. Ved å sette inn $\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ kan (3.21) skrives som:

$$Q = [\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} - \mathbf{m}]' [\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}]^{-1} [\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} - \mathbf{m}]$$

Videre gjelder

$$\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} - \mathbf{m} = \mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'[\mathbf{y} - \mathbf{X}\mathbf{K}(\mathbf{K}'\mathbf{K})^{-1}\mathbf{m}]$$

siden vi vet ved Lemma 3.3 at $\mathbf{K}'\mathbf{K}$ er ikke-singulær (\mathbf{K}' har full rad-rang). Settes dette inn i uttrykket ovenfor fås:

$$Q = [\mathbf{y} - \mathbf{X}\mathbf{K}(\mathbf{K}'\mathbf{K})^{-1}\mathbf{m}]' \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}[\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}]^{-1}\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'[\mathbf{y} - \mathbf{X}\mathbf{K}(\mathbf{K}'\mathbf{K})^{-1}\mathbf{m}]$$

Vi husker at $SSE = \mathbf{y}'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y}$. Dessuten at $\mathbf{X}'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] = [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{X} = \mathbf{0}$. Da kan SSE omskrives til en kvadratisk form i samme uttrykk som Q :

$$SSE = [\mathbf{y} - \mathbf{X}\mathbf{K}(\mathbf{K}'\mathbf{K})^{-1}\mathbf{m}]' [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] [\mathbf{y} - \mathbf{X}\mathbf{K}(\mathbf{K}'\mathbf{K})^{-1}\mathbf{m}]$$

Vi vet fra før at både $\frac{SSE}{\sigma^2}$ og $\frac{Q}{\sigma^2}$ er χ^2 fordelt og det ser vi bekreftelse på her siden begge er kvadratiske former i vektoren $\mathbf{y} - \mathbf{X}\mathbf{K}(\mathbf{K}'\mathbf{K})^{-1}\mathbf{m}$ som er normalfordelt. De to korresponderende matrisene i de kvadratiske formene er idempotente, men enda viktigere er at produktet av dem er 0-matrisen, altså:

$$[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}[\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}]^{-1}\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{0}$$

og ved Teorem 3.5 er da de SSE og Q uavhengige (\mathbf{V} i teoremet er her $\sigma^2\mathbf{I}$). Dermed vil F -observatoren $F(H) = \frac{Q/s}{SSE/(N-r(\mathbf{X}))}$ ha flg. fordeling (fra (3.14) og (3.22)):

$$(3.23) \quad F(H) = \frac{Q/s}{SSE/(N-r(\mathbf{X}))} \sim F' \left[s, N-r(\mathbf{X}), (\mathbf{K}'\hat{\mathbf{b}} - \mathbf{m})' [\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}]^{-1} (\mathbf{K}'\hat{\mathbf{b}} - \mathbf{m}) / 2\sigma^2 \right]$$

altså en ikke-sentral F -fordeling, som blir sentral under vår generelle hypotese i (3.19).

$$(3.24) \quad F(H) = \frac{Q}{s\hat{\sigma}^2} = \frac{(\mathbf{K}\hat{\mathbf{b}} - \mathbf{m})' [\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}]^{-1} (\mathbf{K}\hat{\mathbf{b}} - \mathbf{m})}{s\hat{\sigma}^2} \sim F_{s, N-r(\mathbf{X})}$$

Dette resultatet gjelder altså for alle hypoteser som kan skrives som i (3.19). Siden (3.19) medfører at ikke-sentralitetsparameteren = 0 er dette ekvivalent med at når test-observatoren i (3.24) forkaster at den kommer fra en sentral F -fordeling medfører dette forkastning av (3.19). De fire vanlige tilfellene nevnt i innledningen er bare å sette rett inn i (3.24). F.eks. hvis vi skal teste om en bestemt parameter i \mathbf{b} er signifikant forskjellig fra 0, blir dette en F -test med 1 frihetsgrad som er den vanlige t -testen.

4. Modeller som ikke har full rang

Modellen som betraktes nå kan fremdeles gis på formen i (2.1) og (2.2), altså:

$$(4.1) \quad \mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$$

der \mathbf{y} er en $N \times 1$ vektor med observasjoner y_i , \mathbf{b} er en $p \times 1$ vektor med parametre (i stedet for $k+1$ som i regresjonstilfellet i kapittel 3 hvor det var viktig å skille mellom modellen med og uten intercept og k var antall kovariater), \mathbf{X} er en $N \times p$ kjent matrise (vanligvis med 0-er og 1-ere) og \mathbf{e} er som før en støyvvektor med

$$(4.2) \quad E(\mathbf{e}) = \mathbf{0}, \quad \text{var}(\mathbf{e}) = \sigma^2 \mathbf{I}$$

I eksempel 2.1 så vi en illustrasjon på denne modellen som ikke har full rang på \mathbf{X} , og $\mathbf{X}'\mathbf{X}$. Der så vi på den enkleste typen modell, en 1-veis klassifisering. Den modellen som vil gå igjen i eksempler videre, som er den enkleste modellen som likevel er komplisert nok til å få frem de viktige lineæralgebraiske og statistiske resultatene er en 2-veis kryssklassifisering med (og uten) interaksjon. Den er gitt som:

$$(4.3) \quad y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$$

der modellen av observasjonene kan beskrives ved to forskjellige faktorer og deres interaksjon, der α_i er den ene faktoren med a nivåer slik at $i = 1, \dots, a$, β_j er den andre faktoren med b nivåer slik at $j = 1, \dots, b$, og γ_{ij} er interaksjonen mellom faktorene (som beskriver hvordan den ene faktoren varierer for forskjellige nivåer av den andre faktoren). Det er viktig å legge merke til at for alle nivåer i en faktor er alle nivåer av den andre faktoren definert, man kan tenke seg datastrukturen illustrert ved en tabell med radene som den ene faktoren og kolonnene som den andre. Antall observasjoner i hver rute (en verdi for i og j) er gitt ved n_{ij} slik at $k = 1, \dots, n_{ij}$. Radsummen, kolonnesummen og totalsummen av antall observasjoner er gitt ved: $n_{i.} = \sum_{j=1}^b n_{ij}$, $n_{.j} = \sum_{i=1}^a n_{ij}$ og $N = n_{..} = \sum_i \sum_j n_{ij}$. Som beskrevet i kapittel 3 er dataene balansert hvis alle rutene har samme antall observasjoner, altså at $n_{ij} = n$ for alle i og j . Her blir det betraktet som et spesialtilfelle. Det kan godt mangle observasjoner i en eller flere ruter slik at $n_{ij} = 0$ der.

Et eksempel på modell (4.3) på formen i (4.1) med $a = 3$ (3 nivåer i rad-faktoren) og $b = 2$ (2 nivåer i kolonne-faktoren) blir:

$$(4.3) \quad \begin{pmatrix} y_{111} \\ \vdots \\ y_{11n_{11}} \\ y_{121} \\ \vdots \\ y_{12n_{12}} \\ y_{211} \\ \vdots \\ y_{21n_{21}} \\ y_{221} \\ \vdots \\ y_{22n_{22}} \\ y_{311} \\ \vdots \\ y_{31n_{31}} \\ y_{321} \\ \vdots \\ y_{32n_{32}} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & 1 & 0 & 1 & 0 & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & 0 & 1 & 0 & 1 & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & 1 & 0 & \vdots & 0 & 1 & \vdots & 1 & \vdots & \vdots & \vdots & \vdots \\ \vdots & 0 & 1 & \vdots & 1 & 0 & \vdots & 0 & 1 & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & 1 & 0 & \vdots & \vdots & 1 & 0 & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & 0 & 1 & \vdots & \vdots & 0 & 1 & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & 1 & 0 & 0 & 1 & \vdots & \vdots & \vdots & 1 & 0 & \vdots \\ \vdots & \vdots & 0 & 1 & 1 & 0 & \vdots & \vdots & \vdots & 0 & 1 & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & 1 & 0 & \vdots & \vdots & \vdots & \vdots & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & 0 & 1 & \vdots & \vdots & \vdots & \vdots & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta_1 \\ \beta_2 \\ \gamma_{11} \\ \gamma_{12} \\ \gamma_{21} \\ \gamma_{22} \\ \gamma_{31} \\ \gamma_{32} \end{pmatrix} + \begin{pmatrix} \varepsilon_{111} \\ \vdots \\ \varepsilon_{11n_{11}} \\ \varepsilon_{121} \\ \vdots \\ \varepsilon_{12n_{12}} \\ \varepsilon_{211} \\ \vdots \\ \varepsilon_{21n_{21}} \\ \varepsilon_{221} \\ \vdots \\ \varepsilon_{22n_{22}} \\ \varepsilon_{311} \\ \vdots \\ \varepsilon_{31n_{31}} \\ \varepsilon_{321} \\ \vdots \\ \varepsilon_{32n_{32}} \end{pmatrix}$$

Det er ikke vanskelig å se at \mathbf{X} ikke har full kolonne-rang her. Vi ser at summen av de 3 kolonnene som tilsvarende α -faktoren gir den første kolonnen som tilsvarende μ . Sånn er det også for kolonnene assosiert med β og γ . Det er også verdt å merke seg at rommet utspent av kolonnene assosiert med γ inneholder hele rommet som utspennes av kolonnene assosiert med α eller β , f.eks. sum av to γ kolonner gir en α kolonne (de to første γ kolonnene gir den første α kolonnen) slik at alle vektorer som kan uttrykkes ved α -faktoren kan også uttrykkes ved γ -faktoren.

4.1 Generelt

Som beskrevet i kapittel 2 (og i eksempel 2.1) har normalligningene nå ikke noen unik løsning $\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ siden $\mathbf{X}'\mathbf{X}$ ikke har full rang og dermed ikke er invertibel. Normalligningene som nå har mange løsninger kan skrives

$$(4.4) \quad \mathbf{X}'\mathbf{X}\mathbf{b}^0 = \mathbf{X}'\mathbf{y}$$

der \mathbf{b}^0 brukes som symbol på en av mange mulige løsninger til normalligningene, i motsetning til den unike løsningen $\hat{\mathbf{b}}$ i tilfellet med full rang, som er et estimat for \mathbf{b} . For å kunne finne en av de mange løsningene innfører vi den generaliserte inversen \mathbf{G} til $\mathbf{X}'\mathbf{X}$ og skriver løsningen som:

$$(4.5) \quad \mathbf{b}^0 = \mathbf{G}\mathbf{X}'\mathbf{y}$$

Siden det er mange løsninger (\mathbf{b}^0 som oppfyller (4.5)) betyr det at for hver løsning er det en tilhørende generalisert invers \mathbf{G} . Vi kan altså tenke på \mathbf{G} som en invers, men mer generell. Den formelle definisjonen av den generaliserte inversen er gitt ved (se [3], s 1):

Definisjon 4.1: En generalisert invers til en matrise \mathbf{A} er definert som hvilken som helst matrise \mathbf{G} som oppfyller $\mathbf{AGA} = \mathbf{A}$

Til enhver matrise er det uendelig mange generaliserte inverser som oppfyller ligningen i definisjonen. Det er verdt å merke seg at \mathbf{A} ikke trenger være kvadratisk. Hvis dimensjonen til \mathbf{A} er $p \times q$ er dimensjonen til \mathbf{G} $q \times p$.

Et viktig teorem gjør oss i stand til å uttrykke enhver løsning på formen i (4.5) (se [3], s 26):

Teorem 4.1: For alle mulige generaliserte inverser \mathbf{G} til \mathbf{A} , $\tilde{\mathbf{x}} = \mathbf{G}\mathbf{y}$ genererer alle løsninger til de konsistente ligningene $\mathbf{Ax} = \mathbf{y}$.

Med \mathbf{A} som $\mathbf{X}'\mathbf{X}$, $\tilde{\mathbf{x}}$ som \mathbf{b}^0 og \mathbf{y} som $\mathbf{X}'\mathbf{y}$ i ligning (4.4) har vi at for konsistente normalligninger vil (4.5) for alle generaliserte inverser til $\mathbf{X}'\mathbf{X}$, generere alle løsningene til (4.4).

\mathbf{b}^0 vil omtales som en løsning (til normalligningene) og ikke en estimator. Den vil riktignok kunne betraktes som en estimator for "noe", men ikke for \mathbf{b} . Dette kan ses av forventningen til \mathbf{b}^0 :

$$(4.6) \quad E(\mathbf{b}^0) = \mathbf{GX}'E(\mathbf{y}) = \mathbf{GX}'\mathbf{Xb}$$

Det betyr at \mathbf{b}^0 er en forventningsrett estimator for en lineær funksjon av \mathbf{b} , men ikke \mathbf{b} selv. Variansen til \mathbf{b}^0 er gitt ved:

$$(4.7) \quad \text{var}(\mathbf{b}^0) = \text{var}(\mathbf{GX}'\mathbf{y}) = \mathbf{GX}'\text{var}(\mathbf{y})\mathbf{XG}' = \mathbf{GX}'\mathbf{XG}'\sigma^2$$

Som i kapittel 3 er de estimerte forventede y-verdiene (predikerte y-verdiene) $\hat{\mathbf{y}} = E(\widehat{\mathbf{y}})$, og gitt ved:

$$(4.8) \quad \hat{\mathbf{y}} = E(\widehat{\mathbf{y}}) = \mathbf{Xb}^0 = \mathbf{XGX}'\mathbf{y}$$

Denne formen er viktig fordi her ser man grunnformen av de estimerbare funksjonene. Det viser seg nemlig at uansett hvilken løsning (\mathbf{b}^0) man setter inn i (4.8) får man samme resultat, nettopp egenskapen til de estimerbare funksjonene illustrert i eksempel 2.1.

Et teorem som oppsummerer de fire nyttigste egenskapene til den generaliserte inversen og der denne invariansegenskapen er tydelig er flg.:

Teorem 4.2: Når \mathbf{G} er en generalisert invers til $\mathbf{X}'\mathbf{X}$ da gjelder:

- (i) \mathbf{G}' er også en generalisert invers til $\mathbf{X}'\mathbf{X}$
- (ii) $\mathbf{XGX}'\mathbf{X} = \mathbf{X}$; m.a.o \mathbf{GX}' er en generalisert invers til \mathbf{X}
- (iii) \mathbf{XGX}' er invariant til \mathbf{G}
- (iv) \mathbf{XGX}' er symmetrisk, uansett om \mathbf{G} er, eller ikke

Ved å bruke punkt (i) i teoremet på de andre punktene får man dette korollaret:

Korollar:	(i) $\mathbf{XG}'\mathbf{X}'\mathbf{X} = \mathbf{X}$ (ii) $\mathbf{X}'\mathbf{XG}\mathbf{X}' = \mathbf{X}'$ og $\mathbf{X}'\mathbf{XG}'\mathbf{X}' = \mathbf{X}'$ (iii) $\mathbf{XG}'\mathbf{X}' = \mathbf{XG}\mathbf{X}'$ og $\mathbf{XG}'\mathbf{X}'$ er symmetrisk
------------------	---

Punkt (iii) i teoremet sier nettopp at uansett hvilken \mathbf{G} man bruker i (4.8), eller hvilken \mathbf{b}^0 man bruker får man samme $\hat{\mathbf{y}}$. Det holder altså med hvilken som helst løsning av normalligningene for å finne $\hat{\mathbf{y}}$.

Residualfeilkvadratsummen, SSE er helt tilsvarende tilfellet med full rang og uttrykt ved:

$$(4.9) \quad \begin{aligned} SSE &= (\mathbf{y} - \mathbf{Xb}^0)' (\mathbf{y} - \mathbf{Xb}^0) = \mathbf{y}'(\mathbf{I} - \mathbf{XG}\mathbf{X}')(\mathbf{I} - \mathbf{XG}\mathbf{X}')\mathbf{y} \\ &= \mathbf{y}'(\mathbf{I} - \mathbf{XG}\mathbf{X}')\mathbf{y} = \mathbf{y}'\mathbf{y} - \mathbf{b}^{0'}\mathbf{X}'\mathbf{y} \end{aligned}$$

siden $(\mathbf{I} - \mathbf{XG}\mathbf{X}')$ er symmetrisk (punkt (iv)) og idempotent (punkt (ii)). Vi ser også at siden $\mathbf{XG}\mathbf{X}'$ er invariant til \mathbf{G} er også SSE det, så slik som $\hat{\mathbf{y}}$ blir også SSE den samme for alle \mathbf{G} (eller \mathbf{b}^0). Den siste formen er akkurat som i regresjon (3.5).

Siden $\mathbf{y} \sim (\mathbf{Xb}, \mathbf{I}\sigma^2)$ gjelder som i kapittel 3 at forventningen til SSE er gitt ved Teorem 3.1 som:

$$(4.10) \quad \begin{aligned} E(SSE) &= tr[\mathbf{I} - \mathbf{XG}\mathbf{X}']\mathbf{I}\sigma^2 + \mathbf{b}'\mathbf{X}'[\mathbf{I} - \mathbf{XG}\mathbf{X}']\mathbf{Xb} \\ &= r[\mathbf{I} - \mathbf{XG}\mathbf{X}']\sigma^2 = [N - r(\mathbf{X})]\sigma^2 \end{aligned}$$

Det kvadratiske leddet i \mathbf{b} blir $\mathbf{0}$ pga. punkt (ii) i teoremet. Som i kapittel 3 kan vi bruke Lemma 3.1 til å finne at $r[\mathbf{I} - \mathbf{XG}\mathbf{X}'] = N - r(\mathbf{XG}\mathbf{X}')$. Så kan vi, også tilsvarende som i kapittel 3 bruke Teorem 3.2 noen ganger for å vise at $r(\mathbf{XG}\mathbf{X}') = r(\mathbf{X})$. Ved Teorem 3.2 vet vi direkte at $r(\mathbf{XG}\mathbf{X}') \leq r(\mathbf{X})$.

Dessuten kan vi ved Teorem 4.2, punkt (ii) omskrive \mathbf{X} som $\mathbf{X} = \mathbf{XG}\mathbf{X}'\mathbf{X}$ og da får vi ved Teorem 3.2 at $r(\mathbf{X}) = r(\mathbf{XG}\mathbf{X}'\mathbf{X}) \leq r(\mathbf{XG}\mathbf{X}')$. Altså har vi vist at $r(\mathbf{XG}\mathbf{X}') = r(\mathbf{X})$ og forventningen til SSE er som i (4.10). Dermed har vi også funnet et forventningrett estimat for σ^2 som viser seg å være som i regresjonstilfellet:

$$(4.11) \quad \hat{\sigma}^2 = \frac{SSE}{N - r(\mathbf{X})}$$

bare at her vet vi ikke rangen til \mathbf{X} (ikke lik antall kolonner), så nevneren kan ikke forenkles ytterligere. Rangen til \mathbf{X} er avhengig av dataene og modellen.

Helt tilsvarende regresjonsmodellen deler vi, for å isolere de forskjellige kildene til variasjonen i dataene, opp kvadratsummen av observasjonene. De forskjellige bidragene kan da tabuleres i en ANOVA tabell (variensanalyse-tabell):

$$(4.12) \quad SSR = SST - SSE = \mathbf{b}^{0'}\mathbf{X}'\mathbf{y}$$

er som tidligere den delen av den totale kvadratsummen (SST) som modellen bidrar med (regresjonskvadratsummen) og SSE er som i (4.9). Som i kapittel 3 er

$$(4.13) \quad SSM = N \bar{y}^2$$

bidraget til kvadratsummen ved å tilpasse en konstant (globalt gjennomsnitt) til dataene, og

$$(4.14) \quad SSR_m = SSR - SSM = SSR - N \bar{y}^2$$

er regresjons-kvadratsummen korrigert for gjennomsnittet.

4.2 Fordelingsegenskaper

Som i kapittel 3 gjør vi antagelse om normalfordelt støy,

$$(4.15) \quad \mathbf{e} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$$

\mathbf{y} er fordelt som før:

$$(4.16) \quad \mathbf{y} \sim N(\mathbf{X}\mathbf{b}, \sigma^2 \mathbf{I})$$

Fordelingen til \mathbf{b}^0 blir litt annerledes enn fordelingen til $\hat{\mathbf{b}}$, men den er fremdeles en normalfordeling siden \mathbf{b}^0 er en lineærkombinasjon av \mathbf{y} som er normalfordelt. Vi har:

$$(4.17) \quad \mathbf{b}^0 = \mathbf{GX}'\mathbf{y} \sim N(\mathbf{GX}'\mathbf{X}\mathbf{b}, \mathbf{GX}'\mathbf{X}\mathbf{G}'\sigma^2)$$

fra (4.6) og (4.7). Til forskjell fra regresjonstilfellet er nå variansmatrisen til \mathbf{b}^0 singular, det ser vi siden dimensjonen er p , og rangen må ved Teorem 3.2 være mindre enn p , $r(\mathbf{GX}'\mathbf{X}\mathbf{G}') \leq r(\mathbf{X}'\mathbf{X}) < p$.

Som i kapittel 3 har vi at:

$$(4.18) \quad \mathbf{b}^0 \text{ og } \hat{\sigma}^2 \text{ er uavhengige}$$

Ved å bruke Teorem 3.3 med $\mathbf{b}^0 = \mathbf{GX}'\mathbf{y}$, $SSE = \mathbf{y}'(\mathbf{I} - \mathbf{X}\mathbf{G}\mathbf{X}')\mathbf{y}$ fra (4.9) som den kvadratiske formen i \mathbf{y} og kovariansmatrisen til \mathbf{y} , fås $\mathbf{GX}'\mathbf{I}\sigma^2(\mathbf{I} - \mathbf{X}\mathbf{G}\mathbf{X}') = \mathbf{G}(\mathbf{X}' - \mathbf{X}'\mathbf{X}\mathbf{G}\mathbf{X}')\sigma^2 = \mathbf{0}$ fra punkt (ii) i korollaret til Teorem 4.2. Da er \mathbf{b}^0 og SSE uavhengige, og dermed også \mathbf{b}^0 og $\hat{\sigma}^2$.

Fordelingen til SSE/σ^2 er den samme χ^2 fordelingen som i kapittel 3. Ved å bruke formen $SSE/\sigma^2 = \mathbf{y}'(\mathbf{I} - \mathbf{X}\mathbf{G}\mathbf{X}')\mathbf{y}/\sigma^2$ ser vi at $(\mathbf{I} - \mathbf{X}\mathbf{G}\mathbf{X}')/\sigma^2(\mathbf{I}\sigma^2) = (\mathbf{I} - \mathbf{X}\mathbf{G}\mathbf{X}')$ som er idempotent gir fordelingen til SSE/σ^2 fra Teorem 3.4:

$$(4.19) \quad SSE/\sigma^2 \sim \chi^2[r(\mathbf{I} - \mathbf{X}\mathbf{G}\mathbf{X}'), \mathbf{b}'\mathbf{X}'(\mathbf{I} - \mathbf{X}\mathbf{G}\mathbf{X}')\mathbf{X}\mathbf{b}/2\sigma^2] = \chi^2_{N-r}$$

der som i (4.10) $r(\mathbf{I} - \mathbf{XGX}') = N - r(\mathbf{X}) = N - r$ og ikke-sentralitetsparameteren blir $\mathbf{0}$ fra punkt (ii) i korollaret til Teorem 4.2, så fordelingen er sentral.

Når det gjelder SSR er viser denne seg også å være fordelt som i kapittel 3. Vi skriver den på formen fra (4.12), innsatt for \mathbf{b}^0 , $SSR/\sigma^2 = \mathbf{y}'\mathbf{XGX}'\mathbf{y}/\sigma^2$, der $\mathbf{XGX}'/\sigma^2 (\mathbf{I}\sigma^2)$ er idempotent. Kravet i Teorem 3.3 er oppfylt siden $(\mathbf{XGX}'/\sigma^2)(\mathbf{I}\sigma^2)(\mathbf{I} - \mathbf{XGX}')/\sigma^2 = \mathbf{0}$ som viser uavhengighet mellom SSE/σ^2 og SSE . Fordelingen er ved Teorem 3.4:

$$(4.20) \quad SSR/\sigma^2 \sim \chi^2[r(\mathbf{XGX}'), \mathbf{b}'\mathbf{X}'\mathbf{XGX}'\mathbf{Xb}/2\sigma^2] = \chi^2[r, \mathbf{b}'\mathbf{X}'\mathbf{Xb}/2\sigma^2]$$

der $r(\mathbf{XGX}') = r(\mathbf{X}) = r$ fra (4.10) og forenklingen i ikke-sentralitetsparameteren ses fra punkt (ii) i Teorem 4.2.

Helt tilsvarende er det med SSM . Vi bruker formen $SSM/\sigma^2 = \mathbf{y}'N^{-1}\mathbf{1}\mathbf{1}'\mathbf{y}/\sigma^2$, der vi vet at $N^{-1}\mathbf{1}\mathbf{1}'$ er idempotent, og dermed også $(N^{-1}\mathbf{1}\mathbf{1}'/\sigma^2)(\mathbf{I}\sigma^2)$. M.h.t. uavhengighet av SSE , må vi fra Teorem 3.5 ha at $(N^{-1}\mathbf{1}\mathbf{1}'/\sigma^2)(\mathbf{I}\sigma^2)(\mathbf{I} - \mathbf{XGX}') = N^{-1}(\mathbf{1}\mathbf{1}' - \mathbf{1}\mathbf{1}'\mathbf{XGX}')$ blir $\mathbf{0}$. Alle modellene vi er interessert i her inneholder et globalt gjennomsnitt (μ). Da er alltid første kolonne i \mathbf{X} en kolonne med 1-ere. \mathbf{X} har dimensjon $N \times p$ og kan skrives som:

$$\mathbf{X} = \begin{pmatrix} \mathbf{1} & \mathbf{x}_2 & \cdots & \mathbf{x}_p \end{pmatrix} \text{ og dermed } \mathbf{X}' = \begin{pmatrix} \mathbf{1}' \\ \mathbf{x}_2' \\ \vdots \\ \mathbf{x}_p' \end{pmatrix}$$

Så bruker vi punkt (ii) i korollaret til Teorem 4.2 som sier at $\mathbf{X}'\mathbf{XGX}' = \mathbf{X}'$. Som i kapittel 3 betrakter vi venstresiden av dette som et matriseprodukt av \mathbf{X}' og \mathbf{XGX}' og skriver dette på formen

$$\begin{pmatrix} \mathbf{1}' \\ \mathbf{x}_2' \\ \vdots \\ \mathbf{x}_p' \end{pmatrix} = \mathbf{X}' = \mathbf{X}'\mathbf{XGX}' = \begin{pmatrix} \mathbf{1}' \\ \mathbf{x}_2' \\ \vdots \\ \mathbf{x}_p' \end{pmatrix} \mathbf{XGX}' = \begin{pmatrix} \mathbf{1}'\mathbf{XGX}' \\ \mathbf{x}_2'\mathbf{XGX}' \\ \vdots \\ \mathbf{x}_p'\mathbf{XGX}' \end{pmatrix}$$

Fyller inn for kolonnene i \mathbf{X} bare der det trengs for illustrasjonens skyld. Vi ser at øverste rad i denne ligningen gir oss at $\mathbf{1}' = \mathbf{1}'\mathbf{XGX}'$ gjør at kravet i Teorem 3.5 er oppfylt. Da gir Teorem 3.4 fordelingen til SSM/σ^2 :

$$(4.21) \quad SSM/\sigma^2 \sim \chi^2[r(N^{-1}\mathbf{1}\mathbf{1}'), \mathbf{b}'\mathbf{X}'N^{-1}\mathbf{1}\mathbf{1}'\mathbf{Xb}/2\sigma^2] = \chi^2[1, (\mathbf{1}'\mathbf{Xb})^2/2N\sigma^2]$$

akkurat som i tilfellet med full rang.

Til slutt ser vi på fordelingen til SSR_m som også kan skrives som en kvadratisk form i \mathbf{y} med utgangspunkt i (4.14): $SSR_m = SSR - SSM = \mathbf{y}'(\mathbf{XGX}' - N^{-1}\mathbf{1}\mathbf{1}')\mathbf{y}$. $\mathbf{XGX}' - N^{-1}\mathbf{1}\mathbf{1}'$ er idempotent og gir $\mathbf{0}$ både når ganget med $N^{-1}\mathbf{1}\mathbf{1}'$ (ved samme resonnement som for SSM) og med $(\mathbf{I} - \mathbf{XGX}')$. Som ovenfor gir da Teorem 3.4 og 3.5 at SSR_m er uavhengig av både SSM og SSE , og har flg. fordeling:

$$(4.22) \quad \begin{aligned} \frac{SSR_m}{\sigma^2} &\sim \chi^2 \left[r(\mathbf{XGX}' - N^{-1}\mathbf{1}\mathbf{1}'), \mathbf{b}'\mathbf{X}'(\mathbf{XGX}' - N^{-1}\mathbf{1}\mathbf{1}')\mathbf{Xb} / 2\sigma^2 \right] \\ &= \chi^2 \left[r-1, \mathbf{b}'\mathbf{X}'(\mathbf{I} - N^{-1}\mathbf{1}\mathbf{1}')\mathbf{Xb} / 2\sigma^2 \right] \end{aligned}$$

Forenklingen i ikke-sentralitetsparameteren er som i Teorem 4.2. Rangen til $\mathbf{XGX}' - N^{-1}\mathbf{1}\mathbf{1}'$ må forklares litt. Siden denne matrisen er idempotent, vet vi ([4], side 320) at rangen er lik tracen (summen av diagonalelementene), og vi vet at trace til en matrise er en lineær operator slik at trace til en sum (differanse) av matriser er summen (differansen) av trace til hver av matrisene. Dette er lett å se med 2×2 matrisene \mathbf{A} og \mathbf{B} :

$$tr(\mathbf{A} + \mathbf{B}) = (a_{11} + b_{11}) + (a_{22} + b_{22}) = (a_{11} + a_{22}) + (b_{11} + b_{22}) = tr(\mathbf{A}) + tr(\mathbf{B})$$

Da får vi her: $r(\mathbf{XGX}' - N^{-1}\mathbf{1}\mathbf{1}') = tr(\mathbf{XGX}' - N^{-1}\mathbf{1}\mathbf{1}') = tr(\mathbf{XGX}') - tr(N^{-1}\mathbf{1}\mathbf{1}')$, men $tr(N^{-1}\mathbf{1}\mathbf{1}') = 1$ og vi vet fra før at \mathbf{XGX}' også er idempotent og dermed har trace lik rang, som er $r(\mathbf{XGX}') = r(\mathbf{X}) = r$, slik at vi får at $r(\mathbf{XGX}' - N^{-1}\mathbf{1}\mathbf{1}') = r - 1$, som er frihetsgradene i (4.22). Dette stemmer også med det vi vet om frihetsgrader at man mister en for hver parameter man estimerer i uttrykket, og fra (4.14) har vi at SSR_m er som SSR hvis man trekker fra $N\bar{y}^2$, der \bar{y} er å betrakte som et estimat for μ , altså mister vi en frihetsgrad i forhold til SSR , som vi har sett fra (4.20) har r frihetsgrader.

Som i kapittel 3 (3.18) kan vi nå uttrykke relevante ikke-sentrale F-observatorer, ved å bruke (4.20), (4.21) og (4.22) som tellere og (4.19) som nevner. Da får vi flg. fordelinger:

$$(4.23) \quad \begin{aligned} F(R) &= \frac{SSR/r}{SSE/(N-r)} \sim F' \left[r, N-r, \mathbf{b}'\mathbf{X}'\mathbf{Xb} / 2\sigma^2 \right] \\ F(M) &= \frac{SSM/1}{SSE/(N-r)} \sim F' \left[1, N-r, (\mathbf{1}'\mathbf{Xb})^2 / 2N\sigma^2 \right] \\ F(R_m) &= \frac{SSR_m/(r-1)}{SSE/(N-r)} \sim F' \left[r-1, N-r, \mathbf{b}'\mathbf{X}'(\mathbf{I} - N^{-1}\mathbf{1}\mathbf{1}')\mathbf{Xb} / 2\sigma^2 \right] \end{aligned}$$

og under visse null-hypoteser er disse fordelingene sentrale F-fordelinger som gir oss muligheter for å teste disse hypotesene.

Testobservatorene i (4.23) kan betraktes på flere måter. De er opphav til flere hypoteser som alle er riktige og som gir mer informasjon enn bare å betrakte de på en måte. De enkleste hypotesene fremkommer som noen av de situasjonene som gjør fordelingene i (4.23) sentrale. F.eks. vil null-hypotesen $H: \mathbf{Xb} = \mathbf{0}$ gjøre fordelingen til $F(R)$ til en sentral $F_{r, N-r}$ fordeling som dermed er egnet for å teste H . Det denne hypotesen tester er altså om modellen $E(\mathbf{y}) = \mathbf{Xb}$ "gjør rede for" en signifikant del av variasjonen i \mathbf{y} . Her er det verdt å merke seg at $H: \mathbf{b} = \mathbf{0}$ ikke lar seg teste, selv om denne hypotesen også gir $F(R)$ en sentral F-fordeling. Dette er fordi \mathbf{b} ikke er en estimerbar funksjon (så

lengde den har mer enn ett element, se neste avsnitt). Hvis test-observatoren forkaster å komme fra en sentral F -fordeling forkastes $H: \mathbf{X}\mathbf{b} = \mathbf{0}$ der \mathbf{b} ikke kan skiller ut.

En hypotese knyttet til $F(M)$ som gjør ikke-sentralitetsparameteren til $F(M)$ 0 i (4.23) kan skrives $H: E(\bar{y}) = 0$. Dette ser vi ved å skrive ikke-sentralitetsparameteren som $N[E(\bar{y})]^2 / 2\sigma^2$, så denne er 0 under hypotesen som da gir $F(M)$ en sentral $F_{1, N-r}$ fordeling. En annen hypotese som $F(M)$ tester er $H: \mu = 0$, under modellen $E(y_{ij}) = \mu$ (som i tilfellet med full rang). Det var sånn vi utledet SSM i kapittel 3, og denne hypotesen tester altså om hvorvidt det å tilpasse en modell med kun en konstant (som da estimeres med gjennomsnittet) bidrar med en "signifikant" forklaring av variasjonen i \mathbf{y} . Ikke-sentralitetsparameteren i $F(M)$ blir 0 siden \mathbf{X} nå er en kolonne-vektor og \mathbf{b} har bare ett element og er derfor estimerbar.

$F(R_m)$ tester hypotesen $H: \mathbf{X}\mathbf{b} = \mathbf{0}$ som $F(R)$, men hvor \mathbf{b} nå er en vektor uten μ (det globale gjennomsnittet). Som $F(R)$ tester den om modellen $E(\mathbf{y}) = \mathbf{X}\mathbf{b}$ gjør rede for en stor del av variasjonen i \mathbf{y} , utover modellen $E(\mathbf{y}) = \mu$. Det betyr at når $F(R_m)$ er signifikant er minst en av elementene i \mathbf{b} (som er uten μ) signifikant forskjellig fra 0 (eller en lineærkombinasjon av dem). Videre hvis $F(M)$ først har vist seg signifikant betyr det at modellen med faktorer er signifikant bedre enn modellen $E(\mathbf{y}) = \mu$.

4.3 Estimerbare funksjoner

Vi begynner med definisjonen på estimerbare funksjoner, hentet fra ([3], side 180):

Definisjon 4.2: En lineær funksjon av parametrene i modell (4.1) er estimerbar dersom den er identisk lik en lineær funksjon av forventningen til vektoren av observasjoner, \mathbf{y} , altså at $\mathbf{q}'\mathbf{b}$ er estimerbar (for \mathbf{q}' en gitt radvektor) hvis det finnes en vektor \mathbf{t}' slik at $\mathbf{q}'\mathbf{b} = \mathbf{t}'E(\mathbf{y})$

Her er det viktig å merke seg at vektoren \mathbf{t}' på ingen måte trenger å være unik, det klarer seg at den eksisterer for at $\mathbf{q}'\mathbf{b}$ skal være estimerbar.

Utfra denne definisjonen kan det utledes 5 viktige egenskaper ved de estimerbare funksjonene:

- (i): Forventningen til enhver observasjon er estimerbar.
- (ii): Lineærkombinasjoner av estimerbare funksjoner er estimerbare
- (4.24) (iii): At $\mathbf{q}'\mathbf{b}$ er estimerbar (dette er uavh. av \mathbf{b}) er ekvivalent med at $\mathbf{q}' = \mathbf{t}'\mathbf{X}$ for en \mathbf{t}'
- (iv): Når $\mathbf{q}'\mathbf{b}$ er estimerbar, er $\mathbf{q}'\mathbf{b}^0$ invariant til \mathbf{b}^0 (uavhengig av \mathbf{b}^0)
- (v): $\mathbf{q}'\mathbf{b}^0$ er BLUE ("best linear unbiased estimator") til $\mathbf{q}'\mathbf{b}$

Disse egenskapene er beskrevet i detalj i ([3], side 181), og er stort sett intuitive. Punkt (i) er rett fra definisjonen av estimerbarhet, at $\mathbf{q}'\mathbf{b} = \mathbf{t}'E(\mathbf{y})$ for en vektor \mathbf{t}' . Hvis \mathbf{t}' er en vektor med en 1-er og resten 0-er er $\mathbf{t}'E(\mathbf{y})$ et element av $E(\mathbf{y})$ og en lineær funksjon av parametrene etter definisjonen av modellen og dermed estimerbar. Punkt (ii) følger direkte ved å se f.eks. på summen av to estimerbare

funksjoner og se at den kan skrives på formen $\mathbf{t}'E(\mathbf{y})$. Punkt (iii) kommer av at estimerbarhet kan skrives som $\mathbf{q}'\mathbf{b} = \mathbf{t}'E(\mathbf{y}) = \mathbf{t}'\mathbf{X}\mathbf{b}$ og siden begrepet estimerbarhet ikke har med verdien av \mathbf{b} å gjøre må dette gjelde for alle \mathbf{b} . Da følger punkt (iii) direkte. $\mathbf{t}'\mathbf{X}$ kan betraktes som lineærkombinasjoner av radene i \mathbf{X} . Punkt (iv) har med Teorem (4.1) å gjøre; når $\mathbf{q}'\mathbf{b}$ er estimerbar gir punkt (iii) at $\mathbf{q}'\mathbf{b}^0 = \mathbf{t}'\mathbf{X}\mathbf{b}^0 = \mathbf{t}'\mathbf{X}\mathbf{G}\mathbf{X}'\mathbf{y}$ og denne siste formen er ved Teorem (4.2) invariant til \mathbf{G} . Dermed er $\mathbf{q}'\mathbf{b}^0$ også invariant til \mathbf{G} og ved Teorem (4.1) vet vi at alle mulige forskjellige \mathbf{G} genererer alle de forskjellige løsningene \mathbf{b}^0 til normalligningene (4.4). Altså er $\mathbf{q}'\mathbf{b}^0$ også invariant til \mathbf{b}^0 . Lineariteten i punkt (v) er linearitet i observasjonene og ses fra formen $\mathbf{q}'\mathbf{b}^0 = \mathbf{q}'\mathbf{G}\mathbf{X}'\mathbf{y}$. Forventningsrettetheten ses også direkte ved Teorem 4.2 og (4.24) punkt (iii)

$$(4.25) \quad E(\mathbf{q}'\mathbf{b}^0) = \mathbf{q}'E(\mathbf{b}^0) = \mathbf{q}'\mathbf{G}\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{t}'\mathbf{X}\mathbf{G}\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{t}'\mathbf{X}\mathbf{b} = \mathbf{q}\mathbf{b}$$

Når det gjelder "best-heten" er det mhp varians, at dette er den estimatoren som blant alle lineære og forventningsrette estimatorene har minst varians. Denne variansen er gitt ved:

$$(4.26) \quad \text{var}(\mathbf{q}'\mathbf{b}^0) = \mathbf{q}'\mathbf{G}\mathbf{X}'\mathbf{X}\mathbf{G}'\mathbf{q}\sigma^2 = \mathbf{q}'\mathbf{G}\mathbf{X}'\mathbf{X}\mathbf{G}'\mathbf{X}'\mathbf{t}\sigma^2 = \mathbf{q}'\mathbf{G}\mathbf{X}'\mathbf{t}\sigma^2 = \mathbf{q}'\mathbf{G}\mathbf{q}\sigma^2$$

der $\text{var}(\mathbf{b}^0)$ er hentet fra (4.7) og ellers (4.24) punkt (iii) er brukt. At denne variansen er minst av alle variansene til de forventningsrette lineære estimatorene er lett å vise, men vi tar det ikke med her (se [3], side 182). (4.26) viser at selv om ikke $\text{var}(\mathbf{b}^0)$ i (4.7) er en direkte analog til tilfellet med full rang fra kapittel 3 som er $(\mathbf{X}'\mathbf{X})^{-1}\sigma^2$, blir $\text{var}(\mathbf{q}'\mathbf{b}^0)$ seende ut som om $\text{var}(\mathbf{b}^0)$ hadde vært $\mathbf{G}\sigma^2$, som hadde vært en analog. Det viktige med (4.26) er at denne er minst (blant alternativene) og at den, i likhet med $\mathbf{q}'\mathbf{b}^0$ selv, er invariant til \mathbf{G} for estimerbare funksjoner (når funksjonen er invariant må variansen også være det). Hvis $\mathbf{q}'\mathbf{b}$ er en estimerbar funksjon er $\mathbf{q}'\mathbf{b}^0$ BLUE og med varians $\mathbf{q}'\mathbf{G}\mathbf{q}\sigma^2$ og både estimatet og variansen er lik for alle \mathbf{G} (og alle løsningene \mathbf{b}^0 til normalligningene).

Tre viktige spesialtilfeller av estimerbare funksjoner nevnes spesielt.

Fra definisjonen vet vi at lineærkombinasjoner av $\mathbf{X}\mathbf{b}$ er estimerbare, f.eks. $\mathbf{m}'\mathbf{X}\mathbf{b}$ for en vilkårlig vektor \mathbf{m} . Videre er også lineærkombinasjoner av $\mathbf{X}'\mathbf{X}\mathbf{b}$, f.eks. $\mathbf{s}'\mathbf{X}'\mathbf{X}\mathbf{b}$ siden dette også er en lineærkombinasjon av $\mathbf{X}\mathbf{b}$ (med $\mathbf{m}' = \mathbf{s}'\mathbf{X}'$). Tilsvarende er også lineærkombinasjoner av $E(\mathbf{b}^0) = \mathbf{G}\mathbf{X}'\mathbf{X}\mathbf{b}$ også estimerbare, gitt ved f.eks. $\mathbf{w}'E(\mathbf{b}^0) = \mathbf{w}'\mathbf{G}\mathbf{X}'\mathbf{X}\mathbf{b}$ siden dette også er en lineærkombinasjon av $\mathbf{X}\mathbf{b}$ (med $\mathbf{m}' = \mathbf{w}'\mathbf{G}\mathbf{X}'$).

Oppsummert og med estimatene (BLUE) gir dette:

1. $\mathbf{m}'\mathbf{X}\mathbf{b}$ er estimerbar med estimat $\mathbf{m}'\mathbf{X}\mathbf{b}^0 = \mathbf{m}'\mathbf{X}\mathbf{G}\mathbf{X}'\mathbf{y}$
- (4.27) 2. $\mathbf{s}'\mathbf{X}'\mathbf{X}\mathbf{b}$ er estimerbar med estimat $\mathbf{s}'\mathbf{X}'\mathbf{X}\mathbf{b}^0 = \mathbf{s}'\mathbf{X}'\mathbf{y}$ (Teorem 4.2)
3. $\mathbf{w}'\mathbf{G}\mathbf{X}'\mathbf{X}\mathbf{b}$ er estimerbar med estimat $\mathbf{w}'\mathbf{G}\mathbf{X}'\mathbf{X}\mathbf{b}^0 = \mathbf{w}'\mathbf{G}\mathbf{X}'\mathbf{X}\mathbf{G}\mathbf{X}'\mathbf{y} = \mathbf{w}'\mathbf{G}\mathbf{X}'\mathbf{y} = \mathbf{w}'\mathbf{b}^0$

I punkt 2 ser vi at en lineær funksjon av venstresiden i normalligningene er estimerbar med den samme funksjonen av høyresiden i normalligningene. Den viktigste forskjellen mellom punktene i (4.27) er at dimensjonen på vektorene av vilkårlige elementer er forskjellig, og enkelheten i estimatene er forskjellig. Siden \mathbf{X} har dimensjon $N \times p$ har \mathbf{m}' N elementer, mens \mathbf{s}' og \mathbf{w}' begge har p elementer

(som er mindre enn N) og derfor er enklere i formen. Punkt 3 har det enkleste estimatet siden elementene i den vilkårlige vektoren \mathbf{w}' inngår direkte med elementene i \mathbf{b}^0 . I tillegg har den estimerbare funksjonen i punkt 3 oftest noen rader som er bare 0-er, slik at antall elementer i den estimerbare funksjonen som er forskjellig fra 0 er $r = r(\mathbf{X})$ som er mindre enn p . Da har \mathbf{b}^0 også tilsvarende antall elementer som er 0 slik at antall vilkårlige elementer i punkt 3 ofte er r , og dermed den enkleste formen av de estimerbare funksjonene.

Det er uansett bare $r = r(\mathbf{X})$ lineært uavhengige estimerbare funksjoner. For å se dette tas utgangspunkt i den estimerbare funksjonen $\mathbf{q}'\mathbf{b}$ der $\mathbf{q}' = \mathbf{t}'\mathbf{X}$, for vilkårlig \mathbf{t}' . Med en matrise av full rang $\mathbf{T}'_{N \times N}$ vil $\mathbf{Q}' = \mathbf{T}'\mathbf{X}$ gi N estimerbare funksjoner ved radene i $\mathbf{Q}'\mathbf{b}$. Men $r(\mathbf{Q}') = r(\mathbf{X})$ siden \mathbf{T}' har full rang, og derfor er det bare r lineært uavhengige rader i \mathbf{Q}' og dermed r lineært uavhengige elementer i $\mathbf{Q}'\mathbf{b}$. Alle andre estimerbare funksjoner kan altså genereres av disse.

For å sjekke om en funksjon er estimerbar vet vi at det er tilstrekkelig å finne en \mathbf{t}' slik at $\mathbf{t}'\mathbf{X} = \mathbf{q}'$, men dette er ikke nødvendigvis det enkleste (f.eks. når \mathbf{X} har stor dimensjon). Et enklere kriterium er gitt ved:

$$(4.28) \quad \mathbf{q}'\mathbf{b} \text{ er estimerbar hvis og bare hvis } \mathbf{q}'\mathbf{G}\mathbf{X}'\mathbf{X} = \mathbf{q}'$$

Dettes ses ved at når $\mathbf{q}'\mathbf{b}$ er estimerbar så finnes en \mathbf{t}' slik at $\mathbf{q}' = \mathbf{t}'\mathbf{X}$, og da blir $\mathbf{q}'\mathbf{G}\mathbf{X}'\mathbf{X} = \mathbf{t}'\mathbf{X}\mathbf{G}\mathbf{X}'\mathbf{X} = \mathbf{t}'\mathbf{X} = \mathbf{q}'$ altså er det riktig ene veien. Andre veien er å anta $\mathbf{q}'\mathbf{G}\mathbf{X}'\mathbf{X} = \mathbf{q}'$ som nettopp er formen $\mathbf{q}' = \mathbf{t}'\mathbf{X}$ med $\mathbf{t}' = \mathbf{q}'\mathbf{G}\mathbf{X}'$ som betyr estimerbarhet for $\mathbf{q}'\mathbf{b}$. Hvis kravet i (4.28) er oppfylt vet vi at $\mathbf{q}'\mathbf{b}$ er estimerbar altså at $\mathbf{q}'\mathbf{b} = \mathbf{q}'\mathbf{G}\mathbf{X}'\mathbf{X}\mathbf{b}$ er estimerbar og fra punkt 3 i (4.27) er estimatet $\mathbf{q}'\mathbf{b}^0$.

En siste viktig bemerkning angående estimerbarhet er at \mathbf{b} selv ikke er estimerbar. Hvis man ser på ett enkelt element i \mathbf{b} (\mathbf{q}' bestående av en 1-er og resten 0-er) er dette ikke estimerbart fordi det ikke finnes en \mathbf{t}' slik at $\mathbf{t}'\mathbf{X} = \mathbf{q}'$ (man får 1-ere på flere plasser). Når ingen av enkelt-elementene er estimerbare er heller ikke vektoren som helhet det. Man kan også se dette fra kriteriet (4.28). Med en \mathbf{q}' som nevnt (bare en 1-er og resten 0-er) må $\mathbf{G}\mathbf{X}'\mathbf{X} = \mathbf{I}$ for at kriteriet skal være oppfylt. Men dette er ikke tilfellet i vår modell siden dette gjør \mathbf{G} til en venstre-invers til $\mathbf{X}'\mathbf{X}$, og vi vet at $\mathbf{X}'\mathbf{X}$ ikke har venstre invers når den ikke har full kolonne-rang (se [4], 199).

4.4 Den generelle lineære hypotesen

Som i kapittel 3 er vi interessert i den generelle lineære hypotesen som kan skrives som:

$$(4.29) \quad H: \mathbf{K}'\mathbf{b} = \mathbf{m}$$

der \mathbf{K}' har dimensjon $s \times p$ mens \mathbf{b} og \mathbf{m} er som vanlig $p \times 1$ vektorer. Først må man beskrive hvilke hypoteser som kan testes, det er nemlig ikke alle som lar seg teste, selv om de passer i (4.29). De som kan testes, kaller vi testbare hypoteser. Det viser seg at en hypotese er testbar bare hvis $\mathbf{K}'\mathbf{b}$ er estimerbar, for da er $\mathbf{K}'\mathbf{b}^0$ som estimat invariant til \mathbf{b}^0 . En intuitiv forklaring på dette er at en analog til test-observatoren i full-rang tilfellet ville inneholdt $\mathbf{K}'\mathbf{b}^0 - \mathbf{m}$ som da nødvendigvis må være invariant til \mathbf{b}^0 , som oppnås kun når $\mathbf{K}'\mathbf{b}$ er estimerbar. Denne nødvendigheten kan vises formelt (se [3], side 193).

En testbar hypotese er altså en der $\mathbf{K}'\mathbf{b} \equiv \{\mathbf{k}'_i\mathbf{b}\}$ for $i=1, \dots, s$ er slik at $\mathbf{k}'_i\mathbf{b}$ er estimerbar for alle i .

Da må $\mathbf{k}'_i = \mathbf{t}'_i\mathbf{X}$ for gitt \mathbf{t}'_i og dermed kan vi skrive $\mathbf{K}' = \mathbf{T}'\mathbf{X}$ for en matrise $\mathbf{T}'_{s \times N}$. Videre betraktes, som i kapittel 3, kun konsistente og lineært uavhengige hypoteser (alle andre kan genereres av de) slik at vi krever at \mathbf{K}' har full rad-rang s .

Med $\mathbf{K}'\mathbf{b}$ estimerbar og $\mathbf{K}'\mathbf{b}^0$ som estimat (BLUE) fås flg. forventning og varians:

$$(4.30) \quad \begin{aligned} E(\mathbf{K}'\mathbf{b}^0) &= \mathbf{K}'\mathbf{b} \\ \text{var}(\mathbf{K}'\mathbf{b}^0) &= \mathbf{K}'\mathbf{G}\mathbf{X}'\mathbf{X}\mathbf{G}'\mathbf{K}\sigma^2 = \mathbf{K}'\mathbf{G}\mathbf{X}'\mathbf{X}\mathbf{G}'\mathbf{X}'\mathbf{T}\sigma^2 = \mathbf{K}'\mathbf{G}\mathbf{K}\sigma^2 \end{aligned}$$

der vi har brukt (4.7) og $\mathbf{K}' = \mathbf{T}'\mathbf{X}$. For å bruke dette videre må vi først vise at $\mathbf{K}'\mathbf{G}\mathbf{K}$ er ikke-singulær. Siden $\mathbf{K}'\mathbf{b}$ skal være estimerbar og \mathbf{K}' ha full rad-rang vet vi at matrisen $\mathbf{T}'_{s \times N}$ i $\mathbf{K}' = \mathbf{T}'\mathbf{X}$ må ha full rad-rang ettersom hvis den har mindre rang enn s , må produktet også ha det (Teorem 3.2), mens med større rang enn s , må den ha flere rader og da får \mathbf{K}' også det, og noen rader som er lineært avhengige. Det er imidlertid ikke nok at $\mathbf{T}'_{s \times N}$ har full rad-rang, siden fra Teorem 3.2 produktet av 2 matriser godt kan ha mindre rang enn hver av dem. Men eksistensen av en $\mathbf{T}'_{s \times N}$ som ganget med \mathbf{X} gir en $\mathbf{K}'_{s \times p}$ med full rad-rang er lett å vise. Man konstruerer en $\mathbf{T}'_{N \times N}^*$ med full rang som ganget med \mathbf{X} gir en $\mathbf{K}'_{N \times p}^*$. Siden rangen til \mathbf{X} er r som for øvrig fra diskusjonen etter (4.27) også er maksimalt antall lineært uavhengige estimerbare funksjoner vil rangen til $\mathbf{K}'_{N \times p}^*$ være r (fra lemma 3.5 forandres ikke rangen ved multiplikasjon med ikke-singulær matrise). Siden antall ønskede hypoteser ikke nødvendigvis er det maksimale antallet, $s \leq r$, kan vi alltid plukke ut s av radene i $\mathbf{K}'_{N \times p}^*$ og bruke den tilhørende $\mathbf{T}'_{s \times N}$ som genererer disse radene. Å plukke ut en rad fra en matrise kan gjøres ved å gange med en radvektor fra venstre som består av bare 0-er bortsett fra en 1-er på den plassen som tilsvarer den ønskede raden. Hvis flere slike rader plasseres oppå hverandre til en matrise med 1-erne i forskjellige kolonner (man er interessert i forskjellige rader) har denne matrisen full rad-rang. Hvis vi kaller denne rad-utplukkeren for \mathbf{R} kan vår ønskede $\mathbf{T}'_{s \times N}$ finnes fra:

$$\mathbf{K}'_{s \times p} = \mathbf{R}_{s \times N} \mathbf{K}'_{N \times p}^* = \mathbf{R}_{s \times N} \mathbf{T}'_{N \times N}^* \mathbf{X}$$

hvor vår $\mathbf{T}'_{s \times N}$ altså er $\mathbf{T}'_{s \times N} = \mathbf{R}_{s \times N} \mathbf{T}'_{N \times N}^*$ og der 1-erne i \mathbf{R} er slik plassert at de plukker ut s lineært uavhengige rader. Helt tilsvarende må det finnes en $\mathbf{S}'_{s \times p}$ slik at multiplisert med $\mathbf{X}'\mathbf{X}$ gir en $\mathbf{K}'_{s \times p}$ med full rad-rang. Først legger vi merke til at $\mathbf{S}'\mathbf{X}'\mathbf{X}\mathbf{b}$ genererer estimerbare funksjoner etter definisjonen siden hver rad i resultatet er en lineær funksjon av $\mathbf{X}\mathbf{b}$. Grunnen til å innføre denne er som i (4.27) at den har færre vilkårlige elementer enn $\mathbf{T}'_{s \times N}$. Rang til $\mathbf{X}'\mathbf{X}$ er lik rangen til \mathbf{X} (som i kapittel 3), som lett kan ses fra Teorem 4.2. Formen $\mathbf{X} = \mathbf{X}\mathbf{G}\mathbf{X}'\mathbf{X}$ og Teorem 3.2 gir både at $r(\mathbf{X}) \leq r(\mathbf{X}'\mathbf{X})$ og $r(\mathbf{X}'\mathbf{X}) \leq r(\mathbf{X})$ som altså viser at $r(\mathbf{X}) = r(\mathbf{X}'\mathbf{X})$. At $\mathbf{S}'_{s \times p}$ må ha full rad-rang for at $\mathbf{K}'_{s \times p} = \mathbf{S}'_{s \times p} \mathbf{X}'\mathbf{X}$ skal ha full rad-rang ses direkte (som i ste), men alle $\mathbf{S}'_{s \times p}$ med full rad-rang oppfyller ikke dette. For å vise at det må eksistere en $\mathbf{S}'_{s \times p}$ som gjør det, innfører vi $\mathbf{S}'_{p \times p}^*$ med full rang (som alltid eksisterer). Da vil $\mathbf{K}'_{s \times p}$ med full rad-rang være gitt ved:

$$\mathbf{K}'_{s \times p} = \mathbf{R}_{s \times p} \mathbf{K}'_{p \times p}^* = \mathbf{R}_{s \times p} \mathbf{S}'_{p \times p}^* \mathbf{X}'\mathbf{X}$$

der $\mathbf{R}_{s \times p}$ er den som plukker ut s lineært uavhengige rader i $\mathbf{K}'_{p \times p}$ og $\mathbf{S}'_{s \times p} = \mathbf{R}_{s \times p} \mathbf{S}'_{p \times p}$ er den vi ville vise eksistensen til. Rangem til $\mathbf{S}'_{s \times p} \mathbf{X}'$ må fra Teorem 3.2 være $r(\mathbf{S}' \mathbf{X}') \geq r(\mathbf{K}') = s$, mens den pga. sin dimensjon som er $s \times N$ må være $r(\mathbf{S}' \mathbf{X}') \leq s$ altså har vi at $r(\mathbf{S}' \mathbf{X}') = s$. Da følger direkte at

$$(4.31) \quad \mathbf{K}' \mathbf{G} \mathbf{K} = \mathbf{S}' \mathbf{X}' \mathbf{X} \mathbf{G} \mathbf{X}' \mathbf{X} \mathbf{S} = \mathbf{S}' \mathbf{X}' \mathbf{X} \mathbf{S}$$

siden \mathbf{G} er den generaliserte inversen til $\mathbf{X}' \mathbf{X}$. Når $\mathbf{S}' \mathbf{X}'$ har full rad-rang gir Lemma 3.3 at $\mathbf{S}' \mathbf{X}' \mathbf{X} \mathbf{S}$ er positivt definit og derfor ikke-singulær, altså har vi vist at $\mathbf{K}' \mathbf{G} \mathbf{K}$ er ikke-singulær og dermed vil alltid $(\mathbf{K}' \mathbf{G} \mathbf{K})^{-1}$ eksistere.

Nå kan en test-observator utledes for testbare hypoteser, som i kapittel 3 og avsnitt 4.2 om fordelingssegenskaper. Fordelingen til den bygger som vanlig på antagelsen om normalfordelt støy som i (4.15). Dette gir normalfordelt \mathbf{y} , som i (4.16) og dermed normalfordelt \mathbf{b}^0 som i (4.17). Dermed vil en lineær funksjon av \mathbf{b}^0 også være normalfordelt og en test-observator for hypotesen i (4.29) tilsvarende den i kapittel 3 (3.20) vil ha fordelingen:

$$(4.32) \quad \mathbf{K}' \mathbf{b}^0 - \mathbf{m} \sim N(\mathbf{K}' \mathbf{b} - \mathbf{m}, \mathbf{K}' \mathbf{G} \mathbf{K} \sigma^2)$$

med bruk av (4.30). Ved å lage en kvadratisk form i $\mathbf{K}' \mathbf{b}^0 - \mathbf{m}$, tilsvarende i kapittel 3, har denne fra Teorem (3.4) en χ^2 fordeling. Den kvadratiske formen kaller vi Q , og er gitt ved:

$$(4.33) \quad Q = (\mathbf{K}' \mathbf{b}^0 - \mathbf{m})' (\mathbf{K}' \mathbf{G} \mathbf{K})^{-1} (\mathbf{K}' \mathbf{b}^0 - \mathbf{m})$$

som da har flg. fordeling:

$$(4.34) \quad \frac{Q}{\sigma^2} \sim \chi^2 \left[s, (\mathbf{K}' \mathbf{b} - \mathbf{m})' (\mathbf{K}' \mathbf{G} \mathbf{K})^{-1} (\mathbf{K}' \mathbf{b} - \mathbf{m}) / 2 \sigma^2 \right]$$

For å se uavhengigheten med SSE , må Q og SSE omskrives til samme form. Det gjøres ved å bruke at $\mathbf{m} = \mathbf{K}' \mathbf{K} (\mathbf{K}' \mathbf{K})^{-1} \mathbf{m} = \mathbf{T}' \mathbf{X} \mathbf{K} (\mathbf{K}' \mathbf{K})^{-1} \mathbf{m} = \mathbf{T}' \mathbf{X} \mathbf{G} \mathbf{X}' \mathbf{X} \mathbf{K} (\mathbf{K}' \mathbf{K})^{-1} \mathbf{m} = \mathbf{K}' \mathbf{G} \mathbf{X}' \mathbf{X} \mathbf{K} (\mathbf{K}' \mathbf{K})^{-1} \mathbf{m}$, siden vi vet at $(\mathbf{K}' \mathbf{K})^{-1}$ eksisterer fordi \mathbf{K}' har full rad-rang (Lemma 3.3). Da kan Q i (4.33) skrives som:

$$Q = \left[\mathbf{y} - \mathbf{X} \mathbf{K} (\mathbf{K}' \mathbf{K})^{-1} \mathbf{m} \right]' \mathbf{X} \mathbf{G}' \mathbf{K} (\mathbf{K}' \mathbf{G} \mathbf{K})^{-1} \mathbf{K}' \mathbf{G} \mathbf{X}' \left[\mathbf{y} - \mathbf{X} \mathbf{K} (\mathbf{K}' \mathbf{K})^{-1} \mathbf{m} \right]$$

Fordelen med denne formen er at også SSE kan skrives som en kvadratisk form i $(\mathbf{y} - \mathbf{X} \mathbf{K} (\mathbf{K}' \mathbf{K})^{-1} \mathbf{m})$, med utgangspunkt i (4.9), nemlig som:

$$SSE = \left[\mathbf{y} - \mathbf{X} \mathbf{K} (\mathbf{K}' \mathbf{K})^{-1} \mathbf{m} \right]' (\mathbf{I} - \mathbf{X} \mathbf{G} \mathbf{X}') \left[\mathbf{y} - \mathbf{X} \mathbf{K} (\mathbf{K}' \mathbf{K})^{-1} \mathbf{m} \right]$$

pga. at $\mathbf{X}' (\mathbf{I} - \mathbf{X} \mathbf{G} \mathbf{X}') = \mathbf{0}$. Dette gjør også at kravet i Teorem 3.5 er oppfylt at produktet av matrisene i de to kvadratiske formene Q og SSE blir null (siden $(\mathbf{y} - \mathbf{X} \mathbf{K} (\mathbf{K}' \mathbf{K})^{-1} \mathbf{m})$ har samme variansmatrise som \mathbf{y}) og dermed er Q og SSE uavhengige. Fra (4.19) har vi fordelingen til SSE , og dermed kan vi nå uttrykke fordelingen til testobservatoren for den generelle hypotesen (4.29):

$$(4.35) \quad F(H) = \frac{Q/s}{SSE/(N-r)} \sim F' \left[s, N-r, (\mathbf{K}'\mathbf{b} - \mathbf{m})' (\mathbf{K}'\mathbf{G}\mathbf{K})^{-1} (\mathbf{K}'\mathbf{b} - \mathbf{m}) / 2\sigma^2 \right]$$

og under hypotesen blir denne en sentral F -fordeling:

$$(4.36) \quad F(H) = (\mathbf{K}'\mathbf{b}^0 - \mathbf{m})' (\mathbf{K}'\mathbf{G}\mathbf{K})^{-1} (\mathbf{K}'\mathbf{b}^0 - \mathbf{m}) / s\hat{\sigma}^2 \sim F_{s, N-r}$$

siden ikke-sentralitetsparameteren i (4.35) blir 0. Hvis test-observatoren forkaster å komme fra en sentral F -fordeling forkastes (4.29) (tilsvarende diskusjonen under (3.24)). Som i kapittel 3 er dette resultatet svært generelt og hypoteser som er mer knyttet til parametrene er mulig å teste, i motsetning til de som baseres på fordelingene i (4.23) som er mer knyttet til modellen. Testbarhet av hypotesene er knyttet til estimerbarhet og er den eneste begrensningen på hvilke hypoteser som lar seg teste. Generaliteten er slik at alle resultatene i kapittel 3 (full rang) er spesialtilfeller av resultatene her (uten full rang), og lar seg utlede fra disse.

5. Sammenligning av forskjellige modeller

For å kunne sammenligne "godheten" til forskjellige modeller for et bestemt datasett er det nødvendig med en notasjon for å kunne skille de forskjellige modellene fra hverandre. Både i kapittel 3 og 4 så vi at en dekomponering av kvadratsummen av observasjonene var hensiktsmessig etter ligningen:

$SSE = SST - SSR$, der SSR beskriver hvor mye den totale kvadratsummen (SST) reduseres med når man tilpasser en bestemt modell. Siden SSE er proporsjonal med estimatet for variansen til \mathbf{y} , vil SSR forklare hvor mye av variasjonen i \mathbf{y} den aktuelle modellen gjør rede for. For at det skal fremgå av notasjonen hvilken modell som tilpasses benyttes $R(\)$ i stedet for SSR med de tilpassede faktorene i parantesen og der R står for reduksjon (etter notasjon i [3]).

Etter å ha beskrevet estimerbarhet, er det lett å se at dette kommer til nytte i testing av hypoteser. Vi skal se at de samme testobservatorene som f.eks. i (4.23) ble brukt til generelle hypoteser som testet om forskjellige modeller forklarte en signifikant andel av variasjonen i \mathbf{y} osv. også har sin tolkning som konkrete hypoteser om parametrene i modellene (med begrensningene som kommer fra estimerbarhet). Testobservatoren under den generelle lineære hypotesen (om parametrene) er gitt i (4.29) med fordelingen i (4.35). Vi skal se at denne testobservatoren reduseres til de i (4.23) i bestemte tilfeller og at telleren i F -observatoren er et passende $R(\)$ -ledd.

5.1 Tilpasning av en konstant

Som vi har sett før (3.10) vil tilpasning av modellen $y_i = \mu + \varepsilon_i$ (bare en konstant) gi

$$(5.1) \quad R(\mu) = SSR = SSM = N \bar{y}^2 = n \bar{y}^2$$

Fra diskusjonen under (4.23) husker vi at $F(M) = SSM / SSE / (N-r)$ tester hypotesen $H: E(\bar{y}) = 0$, eller $H: \mu = 0$. Her ser vi at telleren i denne testobservatoren nettopp er $R(\mu)$, det betyr at $R(\mu)$ har en dobbelt betydning, den kan tolkes som telleren i testobservatoren som tester hypotesene nevnt over, og den er kvadratsumreduksjonen til den enkle modellen med kun en konstant. \mathbf{X} er i dette tilfellet en kolonne med 1-ere, så $\mathbf{X}'\mathbf{X} = N$ (full rang) så μ er estimerbar med $\hat{\mu} = \bar{y}$.

5.2 1-veis klassifisering

Hvis man tilpasser modellen $y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$ som kalles 1-veis klassifisering, blir notasjonen for kvadratsumreduksjonen $R(\mu, \alpha)$ som markerer at parametrene i modellen er et globalt gjennomsnitt μ og de forskjellige parametrene knyttet til en faktor α . I eksempel 2.1 så vi på denne modellen. Normalligningene i (2.8) var et eksempel på den mer generelle:

$$(5.2) \quad \mathbf{X}'\mathbf{X}\mathbf{b} = \begin{pmatrix} n & n_1 & n_2 & \cdots & n_a \\ n_1 & n_1 & 0 & \cdots & 0 \\ n_2 & 0 & n_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n_a & 0 & 0 & \cdots & n_a \end{pmatrix} \begin{pmatrix} \mu^0 \\ \alpha_1^0 \\ \alpha_2^0 \\ \vdots \\ \alpha_a^0 \end{pmatrix} = \begin{pmatrix} y_{..} \\ y_{1.} \\ y_{2.} \\ \vdots \\ y_{a.} \end{pmatrix} = \mathbf{X}'\mathbf{y}$$

En lett måte å løse denne på er å sette $\mu^0 = 0$. Når man stryker μ^0 fra ligningssystemet (og dermed første rad og kolonne i $\mathbf{X}'\mathbf{X}$) er de gjenværende ligningene lett å løse. Generelt er dette ofte en enkel måte å finne en \mathbf{G} på; først settes $p - r$ elementer i \mathbf{b}^0 lik 0, som gjør de resterende ligningene lette å løse fra et ligningssystem med full rang og invertering av en ikke-singulær matrise. Så finner man en av de mange mulige \mathbf{G} matrisene ved å utvide denne med 0-er på riktige plasser og slik at man får rett dimensjon. Her får vi

$$(5.3) \quad \mathbf{b}^0 = \begin{pmatrix} \mu^0 \\ \alpha_1^0 \\ \alpha_2^0 \\ \vdots \\ \alpha_a^0 \end{pmatrix} = \begin{pmatrix} 0 \\ \bar{y}_{.1} \\ \bar{y}_{.2} \\ \vdots \\ \bar{y}_{.a} \end{pmatrix}, \quad \mathbf{G} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 1/n_1 & 0 & \cdots & 0 \\ 0 & 0 & 1/n_2 & & \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1/n_a \end{pmatrix}$$

Da blir $R(\mu, \alpha)$ også lett å finne som:

$$(5.4) \quad R(\mu, \alpha) = SSR = \mathbf{b}^0{}' \mathbf{X}'\mathbf{y} = \sum_{i=1}^a \bar{y}_{.i} y_{i.} = \sum_{i=1}^a n_i \bar{y}_{.i}^2 = \sum_{i=1}^a y_{i.}^2 / n_i$$

ev. uttrykt ved en β -faktor med b nivåer ved indeks $j = 1, \dots, b$ som:

$$(5.5) \quad R(\mu, \beta) = \sum_{j=1}^b \bar{y}_{.j} y_{.j} = \sum_{j=1}^b n_j \bar{y}_{.j}^2 = \sum_{j=1}^b y_{.j}^2 / n_j$$

I en 1-veis klassifisering med en α -faktor vil verken μ , eller α_i være estimerbare. Dette er lett å se fra definisjonen (Definisjon 4.2), som sier at lineærkombinasjoner av forventningen til observasjonene ($\mu + \alpha_i$) er estimerbare (estimeres ved $\mu^0 + \alpha_i^0 = \bar{y}_{.i}$ fra (5.3)). Dvs. at hvis μ , eller α_i skulle vært estimerbare måtte en lineærkombinasjon av $\mu + \alpha_i$ være lik μ ($\mathbf{q}'\mathbf{b} = \mu = \mathbf{t}'E(\mathbf{y})$), eller α_i , noe som er umulig. Fra (5.2) ser vi at rangen til $\mathbf{X}'\mathbf{X}$ (og til \mathbf{X}) er a , (f.eks. er den øverste raden lik summen av de nedenfor), og da vet vi at det finnes maksimalt a lineært uavhengige estimerbare funksjoner, at alle andre genereres av disse. Det betyr at alle estimerbare funksjoner kan genereres av $\sum_{i=1}^a \lambda_i (\mu + \alpha_i)$ for gitte koeffisienter $\lambda_1, \dots, \lambda_a$. Av dette ser vi at med $\sum_i \lambda_i = 0$ (s.a. $\sum_i \lambda_i \mu = 0$) er $\sum_i \lambda_i \alpha_i$

estimerbar. Et eksempel er differanser mellom to α -er, $\alpha_i - \alpha_k, i \neq k$, der $\lambda_i = 1$ og $\lambda_k = -1$, og resten av koeffisientene er 0.

De mest interessante hypotesene om parametrene i denne modellen skal vi se kan testes ved testobservatorene i (4.23) akkurat som i den enklere modellen med bare et gjennomsnitt. En test basert på $F(M)$ er fremdeles $H: E(\bar{y}) = 0$, men nå er $E(\bar{y}) = \mu / N + \sum_i n_i \alpha_i / N$ så hypotesen lar seg også skrive som $H: N\mu + \sum_i n_i \alpha_i = 0$. At dette er en testbar hypotese ser vi av at uttrykket er en estimerbar funksjon (siden summen av koeffisientene foran α_i -ene er lik koeffisienten foran μ , se det generelle uttrykket). Da lar denne hypotesen seg formulere som den generelle i (4.29) ved å sette: $\mathbf{K}' = (N \quad n_1 \quad \dots \quad n_a)$ og $m = 0$. For å finne Q fra (4.36) brukes fra (5.3): $\mathbf{K}'\mathbf{b}^0 = \sum_i n_i \bar{y}_i = N \bar{y}_..$ og $\mathbf{K}'\mathbf{G}\mathbf{K} = \sum_i n_i = N$ som gir: $Q = \mathbf{b}^0 \mathbf{K} (\mathbf{K}'\mathbf{G}\mathbf{K})^{-1} \mathbf{K}'\mathbf{b}^0 = N \bar{y}_..^2 = SSM$. Dessuten er $s = 1$, så den generelle testobservatoren fra (4.35) reduseres til $SSM / (SSE / (N-r))$ som nettopp er $F(M)$ fra (4.23). Dette er altså en bekreftelse på at $F(M)$ tester hypotesen $H: N\mu + \sum_i n_i \alpha_i = 0$. I tillegg tester den fremdeles $H: \mu = 0$ når man ser bort fra α_i -ene.

En test basert på $F(R_m)$ lar seg utlede på samme måte. Vi skal se at dette er hypotesen om at alle α_i -ene er like, altså: $H: \alpha_1 = \alpha_2 = \dots = \alpha_a$. Ved først å beskrive hypotesen som

$H: \alpha_1 - \alpha_2 = \alpha_1 - \alpha_3 = \dots = \alpha_1 - \alpha_a = 0$ gir dette en $\mathbf{K}' = (\mathbf{0} \quad \mathbf{1}_{a-1} \quad -\mathbf{I}_{a-1})$. Da lar det seg vise på tilsvarende måte som ovenfor med \mathbf{G} fra (5.3) at den generelle Q fra (4.35) reduseres til: $Q = \sum_i n_i \bar{y}_i^2 - N \bar{y}_..^2 = SSR - SSM = SSR_m$ med SSM fra ovenfor, SSR fra (5.4) og SSR_m fra (4.14).

Siden vi ser fra (5.2) at rangen til \mathbf{X} er a , får vi at testobservatoren i (4.35) reduseres til $SSR_m / (r-1) / (SSE / (N-r))$ som nettopp er testobservatoren $F(R_m)$ i (4.23). Vi har altså sett at $F(R_m)$ tester om α_i -ene er like. Likevel gjelder fremdeles at den tester hvorvidt modellen som omfatter en α -effekt utover bare et globalt gjennomsnitt μ forklarer en signifikant andel av variasjonen i \mathbf{y} , slik vi så i (4.23).

Heretter skal vi omtale SSR_m i denne modellen som $R(\alpha|\mu)$, som altså beskriver forskjellen i kvadratsumreduksjon mellom en modell med både μ og α -effekt ($R(\mu, \alpha)$) og en modell med bare μ ($R(\mu)$) (hvor mye mer kvadratsummen reduseres ved å ta med en faktor). Denne er gitt ved:

$$(5.6) \quad R(\alpha|\mu) = R(\mu, \alpha) - R(\mu) = SSR_m = \sum_{i=1}^a n_i \bar{y}_i^2 - n \cdot \bar{y}_..^2 = \sum_{i=1}^a n_i (\bar{y}_i - \bar{y}_..)^2$$

(der siste omformingen ses ved direkte utregning av høyresiden), eventuelt med en β -faktor:

$$(5.7) \quad R(\beta|\mu) = R(\mu, \beta) - R(\mu) = SSR_m = \sum_{j=1}^b n_j \bar{y}_{.j}^2 - n \cdot \bar{y}_..^2 = \sum_{j=1}^b n_j (\bar{y}_{.j} - \bar{y}_..)^2$$

Som ved den enkle modellen i avsnitt 5.1, har også $R(\alpha|\mu)$ en dobbelt betydning. Det er telleren i testobservatoren $F(R_m)$ som tester om α_i -ene er like, og det er økningen i kvadratsumreduksjonen som skyldes en tilpasning av både μ og α fremfor bare μ .

5.3 2-veis kryssklassifisering, uten interaksjon

Hvis man utvider modellen ovenfor til å ha med 2 faktorer, både α og β til $y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}$ har vi en 2-veis kryssklassifisering uten interaksjonsledd. For enkelhets skyld skal vi kun la det være 1, eller ingen observasjon for hver kombinasjon av i og j slik at indeksen k kan sløyfes og modellen skrives som $y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$, uten at dette har noe å si for resultatene. n_{ij} er dermed enten 1 eller 0. Denne modellen er den som har mest komplisert løsning av normalligningene og skal vi se har konsekvenser for andre resultater så derfor tas denne løsningen med. Modellen ser ut som i (4.3) ved å droppe γ -leddene og tilhørende kolonner i \mathbf{X} matrisen. Normalligningene vil ha flg. utseende:

$$(5.8) \quad \mathbf{X}'\mathbf{X}\mathbf{b} = \begin{pmatrix} n_{..} & n_{1.} & \cdots & n_{a.} & n_{.1} & \cdots & n_{.b} \\ n_{1.} & n_{11} & & \mathbf{0} & & & \\ \vdots & & \ddots & & \{n_{ij}\} & & \\ n_{a.} & \mathbf{0} & & n_{a.} & & & \\ n_{.1} & & & & n_{.1} & & \mathbf{0} \\ \vdots & & \{n_{ji}\} & & & \ddots & \\ n_{.b} & & & \mathbf{0} & & & n_{.b} \end{pmatrix} \begin{pmatrix} \mu^0 \\ \alpha_1^0 \\ \vdots \\ \alpha_a^0 \\ \beta_1^0 \\ \vdots \\ \beta_b^0 \end{pmatrix} = \begin{pmatrix} y_{..} \\ y_{1.} \\ \vdots \\ y_{a.} \\ y_{.1} \\ \vdots \\ y_{.b} \end{pmatrix} = \mathbf{X}'\mathbf{y}$$

En måte å finne en løsning på er tilsvarende som for modellen i forrige avsnitt. Etter litt stirring på $\mathbf{X}'\mathbf{X}$ kan man se at hvis dimensjonen er $a + b + 1$ er rangen to mindre (man kan finne de to avhengighetene ved å summere α -radene og få μ -raden og det samme ved summen av β -radene).

Ved oppskriften fra forrige avsnitt settes to elementer i \mathbf{b}^0 til 0 og de tilhørende radene fjernes, for å få en matrise med full rang som kan inverteres. Hvis det er færre β -nivåer enn α -nivåer fjernes raden som tilsvarende β_b (og $\beta_b^0 = 0$) og raden som tilsvarende μ (og $\mu^0 = 0$). Da er α -ligningene uttrykt ved β -ene og løsningen uttrykkes ved β -ene som:

$$\mathbf{C}\boldsymbol{\beta}_{b-1}^0 = \mathbf{r} \text{ med løsning } \boldsymbol{\beta}_{b-1}^0 = \mathbf{C}^{-1} \mathbf{r}$$

der $b - 1$ er dimensjonen på den ukjente β -vektoren, \mathbf{C} en matrise med elementer $\{c_{jj'}\}$ og \mathbf{r} er en vektor med elementer $\{r_j\}$ der $j, j' = 1, \dots, b - 1$ og elementene er gitt ved:

$$(5.9) \quad c_{jj} = n_{.j} - \sum_{i=1}^a \frac{n_{ij}^2}{n_{i.}}, \quad c_{jj'} = -\sum_{i=1}^a \frac{n_{ij}n_{ij'}}{n_{i.}}, \quad j \neq j'$$

og

$$(5.10) \quad r_j = y_{.j} - \sum_{i=1}^a n_{ij} \bar{y}_i, \quad j = 1, \dots, b - 1$$

Løsningene for α -ene kan da finnes ved å sette tilbake i (5.8), og den tilhørende \mathbf{G} finnes ved å invertere matrisen i (5.8) når man tar bort første rad og kolonne og deretter fyller på med passe mengde 0-er. Vi trenger i tillegg matrisene $\mathbf{M}_{a \times (b-1)} = \left\{ \frac{n_{ij}}{n_{i.}} \right\}$ og $\mathbf{D}_a = \text{diag}\{n_{i.}\}$, for $i = 1, \dots, a$ og $j = 1, \dots, b - 1$, og vektorene $\bar{\mathbf{y}}_a = \{\bar{y}_{i.j}\}$, $\mathbf{y}_a = \{y_{i.}\}$ med dimensjon a , og $i = 1, \dots, a$. Da kan løsningen uttrykkes ved:

$$(5.11) \quad \mathbf{b}^0 = \begin{pmatrix} 0 \\ \boldsymbol{\alpha}^0 \\ \boldsymbol{\beta}_{b-1}^0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ \bar{\mathbf{y}}_a - \mathbf{M}\mathbf{C}^{-1}\mathbf{r} \\ \mathbf{C}^{-1}\mathbf{r} \\ 0 \end{pmatrix}, \quad \mathbf{G} = \begin{pmatrix} 0 & \mathbf{0} & \mathbf{0} & 0 \\ \mathbf{0} & \mathbf{D}_a^{-1} + \mathbf{M}\mathbf{C}^{-1}\mathbf{M}' & -\mathbf{M}\mathbf{C}^{-1} & \mathbf{0} \\ \mathbf{0} & -\mathbf{C}^{-1}\mathbf{M}' & \mathbf{C}^{-1} & \mathbf{0} \\ 0 & \mathbf{0} & \mathbf{0} & 0 \end{pmatrix}$$

Da finner vi lett $R(\mu, \alpha, \beta) = \mathbf{b}^{0'} \mathbf{X}' \mathbf{y}$ som i forrige modell. Vi definerer vektoren $\mathbf{y}_{b-1} = \{y_{.j}\}$ for $j=1, \dots, b-1$, tilsvarende som \mathbf{y}_a . Da får vi:

$$(5.12) \quad \begin{aligned} R(\mu, \alpha, \beta) &= (\bar{\mathbf{y}}_a - \mathbf{M}\mathbf{C}^{-1}\mathbf{r})' \mathbf{y}_a + (\mathbf{C}^{-1}\mathbf{r})' \mathbf{y}_{b-1} \\ &= \dots = R(\mu, \alpha) + \mathbf{r}'\mathbf{C}^{-1}\mathbf{r} = \sum_{i=1}^a n_i \bar{y}_i^2 + \boldsymbol{\beta}^{0'} \mathbf{r} \end{aligned}$$

hvor vi har brukt at \mathbf{r} kan skrives på formen $\mathbf{r} = \mathbf{y}_{b-1} - \mathbf{M}'\mathbf{y}_a$ og $R(\mu, \alpha)$ er som i forrige avsnitt ($R(\mu, \alpha) = \sum_{i=1}^a n_i \bar{y}_i^2 = \bar{\mathbf{y}}_a' \mathbf{y}_a$) og har samme tolkningen som før, nemlig kvadratsumreduksjon til modellen uten β -faktor. Dette gjelder også for $R(\mu) = n \bar{y}_{..}^2$ som er tilpassningen uten α og β -faktor.

En kvadratsum som er spesiell for denne modellen er $R(\alpha, \beta | \mu)$ som er hvor mye mer kvadratsummen minker når man tilpasser en α - og β -faktor i tillegg til en μ . Etter samme mønster som i forrige avsnitt, og direkte innsetting er denne gitt ved:

$$(5.13) \quad R(\alpha, \beta | \mu) = R(\mu, \alpha, \beta) - R(\mu) = SSR_m = \sum_{i=1}^a n_i \bar{y}_i^2 + \mathbf{r}'\mathbf{C}^{-1}\mathbf{r} - n \bar{y}_{..}^2$$

Nå er $R(\alpha, \beta | \mu) = SSR_m$, $R(\mu, \alpha, \beta) = SSR$ og $R(\mu) = SSM$. Dette betyr at hypotesen basert på $F(M)$ fra (4.23) og $R(\mu)$ fremdeles tester $H: E(\bar{y}) = 0$, med $E(\bar{y}_{..}) = \left(\mu + \frac{b}{N} \sum_i \alpha_i + \frac{a}{N} \sum_j \beta_j \right)$ mens hypotesen basert på $F(R_m)$ og SSR_m ($R(\alpha, \beta | \mu)$) tester om det er grunnlag for noe mer i modellen enn bare en μ , her i form av faktorene α og β . Hvis hypotesen om at disse er lik 0 forkastes, vet vi ikke om det er den ene, eller den andre, eller begge faktorene som gir opphav til forklaringsevnen og må teste mer. Da kan man begynne med å teste radene (α) alene (utelate β -faktoren), slik at man får en vanlig 1-veis klassifisering som i forrige avsnitt og basere testingen på $R(\alpha | \mu)$ og $R(\mu)$ som der. Men etter det er det behov for å teste β -faktoren etter først å ha testet for μ og α , altså om det er noe gevinst å ta med β når man har de andre parametrene. Denne testen baseres på $R(\beta | \mu, \alpha)$ og er gitt ved:

$$(5.14) \quad R(\beta | \mu, \alpha) = R(\mu, \alpha, \beta) - R(\mu, \alpha) = \boldsymbol{\beta}^{0'} \mathbf{r} = \mathbf{r}'\mathbf{C}^{-1}\mathbf{r}$$

når man setter inn fra (5.4) og (5.12). Summen av $R(\alpha | \mu)$ og $R(\beta | \mu, \alpha)$ gir $R(\alpha, \beta | \mu)$ (ses lett ved innsetting) så dette er en ytterligere dekomponering av (5.13) som den er brukt i en ANOVA tabell.

For å dekke alle mulighetene må dette også gjøres med å teste kolonnene (β) først (utelate α -faktoren) og helt samme fremgangsmåte.

Alle $R(\mathbf{b})$ ledd (med \mathbf{b} som vektoren av de forskjellige parametrene) er *SSR* til en eller annen modell.

Den kan altså uttrykkes som $\mathbf{y}'\mathbf{X}\mathbf{G}\mathbf{X}'\mathbf{y}$ med matrisen i midten idempotent. Med $\mathbf{y} \sim \mathbf{N}(\boldsymbol{\mu}, \sigma^2\mathbf{I})$ for vilkårlig $\boldsymbol{\mu}$ vektor vet vi fra Teorem 3.4 at $R(\mathbf{b})/\sigma^2$ da vil ha en ikke-sentral χ^2 fordeling uavhengig av *SSE*. Hvis man så dekomponerer parameter-vektoren i $\mathbf{b}' = (\mathbf{b}'_1, \mathbf{b}'_2)$ gjelder

$R(\mathbf{b}_1, \mathbf{b}_2) = R(\mathbf{b}_2|\mathbf{b}_1) + R(\mathbf{b}_1)$. Da kan det vises at også $R(\mathbf{b}_2|\mathbf{b}_1)/\sigma^2$ har en ikke-sentral χ^2 fordeling (i tillegg til at $R(\mathbf{b}_1)/\sigma^2$ har) som er uavhengig av $R(\mathbf{b}_1)$ og *SSE* (se [3], s 247). Dermed har vi en generell regel som bekrefter det vi så ovenfor at $R(\alpha, \beta|\mu)$ som utgjør telleren i $F(R_m)$ er ikke-sentralt χ^2 -fordelt. Det kan vi se direkte av formen $R(\mu, \alpha, \beta) = R(\alpha, \beta|\mu) + R(\mu)$. Dette kan vi så bruke videre, ved å skrive $R(\mu, \alpha, \beta) = R(\beta|\mu, \alpha) + R(\mu, \alpha)$ og $R(\mu, \alpha) = R(\alpha|\mu) + R(\mu)$. Da gir samme resonnement at $R(\beta|\mu, \alpha)$ og $R(\alpha|\mu)$ begge er har ikke-sentrale χ^2 -fordelinger (uavhengige av hverandre og *SSE*). Dette og det at disse to leddene summert gir $R(\alpha, \beta|\mu)$ er det som skal til for å kunne dekomponere kvadratsummen i forskjellige deler (som i en ANOVA tabell), som hver representerer en modell og har en tilhørende *F*-observator som lar seg teste for å se om denne modellen representerer en signifikant andel av forklaringen til \mathbf{y} . F.ek. representerer $R(\beta|\mu, \alpha)$ modellen som inkluderer β etter å allerede ha μ og α . Hvis den tilhørende *F*-observatoren med $R(\beta|\mu, \alpha)$ som teller er signifikant betyr altså det at det er en signifikant forbedring å inkludere β .

Til slutt får man mange kvadratsummer med tilhørende *F*-observatorer som kan være signifikante eller ikke. Konklusjonene må trekkes på bakgrunn av alle testene, og er ikke nødvendigvis entydige alltid, m.h.p. hvilke faktorer som skal inkluderes i modellen. Dette er viktig å være klar over.

I denne modellen er verken μ , individuelle α_i -er eller individuelle β_j -er estimerbare, av samme grunn som i forrige avsnitt. Forventet verdi for en observasjon er $E(y_{ij}) = \mu + \alpha_i + \beta_j$, så disse, eller lineær-kombinasjoner av disse er estimerbare, noe som fører til at f.eks. differanser mellom to α_i -er eller to β_j -er er estimerbare, med $\alpha_i^0 - \alpha_h^0$, eller $\beta_j^0 - \beta_k^0$ som estimater (BLUE).

Ved siden av å finne hvilken modell som er best, vha de forskjellige *F*-testene, kan de forskjellige $R(\)$ leddene også brukes til å teste bestemte hypoteser om parametrene i modellen, når modellen er gitt. Nå skal vi se på noen hypoteser som lett kan testes her, altså når modellen er en 2-veis kryssklassifisering uten interaksjon.

For å teste hypotesen om at alle β_j -ene er like altså: $H: \beta_1 = \beta_2 = \dots = \beta_b$ omformuleres denne først til $H: \beta_1 - \beta_b = \dots = \beta_{b-1} - \beta_b = 0$ (slik som i forrige avsnitt), som gjør at vi vet at hypotesene er testbare (siden vi nettopp fant ut at differanser som disse er estimerbare). Da kan denne settes på den generelle formen fra (4.29) med $H: \mathbf{K}'\mathbf{b} = \mathbf{0}$ med $\mathbf{K}' = (\mathbf{0} \quad \mathbf{0} \quad \mathbf{I}_{b-1} \quad -\mathbf{1}_{b-1})$. Videre med \mathbf{G} fra (5.11) får vi $\mathbf{K}'\mathbf{G} = (\mathbf{0} \quad -\mathbf{C}^{-1}\mathbf{M}' \quad \mathbf{C}^{-1} \quad \mathbf{0})$ og dermed $\mathbf{K}'\mathbf{G}\mathbf{K} = \mathbf{C}^{-1}$. I tillegg blir

$\mathbf{K}'\mathbf{b}^0 = -\mathbf{C}^{-1}\mathbf{M}'\mathbf{y}_a + \mathbf{C}^{-1}\mathbf{y}_{b-1}$. Da blir telleren i den generelle test-observatoren fra (4.35) lik:

$$\begin{aligned}
Q &= (\mathbf{K}'\mathbf{b}^0)' (\mathbf{K}'\mathbf{G}\mathbf{K})^{-1} \mathbf{K}'\mathbf{b}^0 \\
&= (-\mathbf{C}^{-1}\mathbf{M}'\mathbf{y}_a + \mathbf{C}^{-1}\mathbf{y}_{b-1})' (\mathbf{C}^{-1})^{-1} (-\mathbf{C}^{-1}\mathbf{M}'\mathbf{y}_a + \mathbf{C}^{-1}\mathbf{y}_{b-1}) \\
&= \mathbf{r}'\mathbf{C}^{-1}\mathbf{r} = \boldsymbol{\beta}'\mathbf{r} = R(\boldsymbol{\beta}|\boldsymbol{\mu}, \boldsymbol{\alpha})
\end{aligned}$$

ved (5.14). Dette betyr at $R(\boldsymbol{\beta}|\boldsymbol{\mu}, \boldsymbol{\alpha})$ kan brukes for å teste hypotesen om at alle β_j -ene er like (som teller i F -observatoren). Tilsvarende vil $R(\boldsymbol{\alpha}|\boldsymbol{\mu}, \boldsymbol{\beta})$ være telleren i test-observatoren for å teste om alle α_i -ene er like.

Her er det viktig å legge merke til at hypotesen som tester likhet blant β_j -ene ikke er basert på $R(\boldsymbol{\beta}|\boldsymbol{\mu})$ som i 1-veis klassifiseringen, men på $R(\boldsymbol{\beta}|\boldsymbol{\mu}, \boldsymbol{\alpha})$. Det kan vises på tilsvarende måte som ovenfor at i denne modellen (med begge faktorene til stede) tester $R(\boldsymbol{\beta}|\boldsymbol{\mu})$ (som teller i F -observatoren) hypotesen om at alle $\beta_j + \frac{1}{n_j} \sum_{i=1}^a n_{ij} \alpha_i$ for $j=1, \dots, b$ er like (se [3] s 282). Tolkningen av denne er likhet blant β_j + et veiet kolonnegjennomsnitt av α_i -ene. For $R(\boldsymbol{\alpha}|\boldsymbol{\mu})$ er det tilsvarende, den er teller i observatoren som tester $H: \alpha_i + \frac{1}{n_i} \sum_{j=1}^b n_{ij} \beta_j$ er like, for alle i .

5.4 2-veis kryss klassifisering, med interaksjon

Dette er modellen som vi brukte som illustrasjon i avsnitt 4.3 og som er den enkleste modellen som får frem alle de interessante kvadratsummene. Den skrives $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$. Som i forrige avsnitt har den n_{ij} observasjoner for hver kombinasjon av i og j . Det er nødvendig med noe

tilleggsnotasjon pga. at vi betrakter tilfellet der n_{ij} ikke lengre er 0 eller 1. Vi lar $y_{ij.} = \sum_{k=1}^{n_{ij}} y_{ijk}$,

$\bar{y}_{ij.} = y_{ij.}/n_{ij}$, $y_{i..} = \sum_{j=1}^b y_{ij.}$, $y_{.j.} = \sum_{i=1}^a y_{ij.}$, med $n_{i.} = \sum_{j=1}^b n_{ij}$ og $n_{.j} = \sum_{i=1}^a n_{ij}$ som i forrige avsnitt.

Datastrukturen kan betraktes som å komme fra en tabell med rader og kolonner. Siden det kan være 0 observasjoner i en rute vil antall nivåer i γ faktoren være lik antall ruter med observasjoner i , altså antall ruter der $n_{ij} > 0$. Vi kaller dette antallet for s . Som i forrige avsnitt er antall nivåer for α er a ,

og for β er b . Normalligningene er i dette tilfellet enklere å løse enn i forrige avsnitt, den ligner på tilfellet med 1-veis klassifisering. Rangens til $\mathbf{X}'\mathbf{X}$ er s , som ikke er vanskelig å se, når vi husker at rommet utspent av kolonnene (el. radene) til α ligger inni kolonnerommet til γ , så også for β . Som

før setter vi $p-r=1+a+b$ elementer i \mathbf{b}^0 lik 0, og velger for enklest mulig løsning $\boldsymbol{\mu}^0 = 0$,

$\alpha_1^0 = \dots = \alpha_a^0 = 0$ og $\beta_1^0 = \dots = \beta_b^0 = 0$. Dette gir den enkle løsningen $\gamma_{ij}^0 = \bar{y}_{ij.}$ for de rutene som har

$n_{ij} \neq 0$ (s stykker). Vi får flg. \mathbf{b}^0 og \mathbf{G} :

$$(5.15) \quad \mathbf{b}^0 = \begin{pmatrix} \mathbf{0}_{1+a+b} \\ \bar{y}_{11.} \\ \vdots \\ \bar{y}_{ab.} \end{pmatrix}, \quad \mathbf{G} = \begin{pmatrix} \mathbf{0}_{(1+a+b) \times (1+a+b)} & \mathbf{0}_{(1+a+b) \times s} \\ \mathbf{0}_{s \times (1+a+b)} & \text{diag}_{s \times s} \left\{ \frac{1}{n_{ij}} \right\} \end{pmatrix}$$

Utrekning av $R(\mu)$, $R(\mu, \alpha)$, $R(\mu, \beta)$, $R(\mu, \alpha, \beta)$ blir som i forrige avsnitt, bortsett fra at $y_{..}, y_{i.}, y_{.j}, y_{ij}$ byttes ut med $\bar{y}_{..}, \bar{y}_{i.}, \bar{y}_{.j}, \bar{y}_{ij}$. Den nye kvadratsummen $R(\mu, \alpha, \beta, \gamma)$ finnes på vanlig måte ved $\mathbf{b}'\mathbf{X}'\mathbf{y}$. Ved å registrere at $\mathbf{X}'\mathbf{y}$ (høyresiden i normalligningene) er en kolonnevektor med y_{ij} totalene, er det lett å se at

$$(5.16) \quad R(\mu, \alpha, \beta, \gamma) = \mathbf{b}'\mathbf{X}'\mathbf{y} = \sum_{i=1}^a \sum_{j=1}^b \bar{y}_{ij} \cdot y_{ij} = \sum_{i=1}^a \sum_{j=1}^b n_{ij} \bar{y}_{ij}^2$$

og med de omtalte utskiftningene har vi også:

$$(5.17) \quad \begin{aligned} R(\mu) &= n_{..} \bar{y}_{..}^2 \\ R(\mu, \alpha) &= \sum_{i=1}^a n_{i.} \bar{y}_{i.}^2 \\ R(\mu, \beta) &= \sum_{j=1}^b n_{.j} \bar{y}_{.j}^2 \\ R(\mu, \alpha, \beta) &= \sum_{i=1}^a n_{i.} \bar{y}_{i.}^2 + \mathbf{r}'\mathbf{C}^{-1}\mathbf{r} \end{aligned}$$

der den nederste av disse tilsvarer en 2-veis kryssklassifisering uten interaksjon (som i forrige avsnitt), men med flere enn 1 observasjon pr. rute. Så alle kvadratsummene (5.17) beregnes på samme måte, enten de er av interesse som en del av en variansanalyse for en 2-veis kryssklassifisering med interaksjon, eller som tilpassning av en mindre modell.

Som i forrige avsnitt er $R(\mu)$ telleren i $F(M)$ som tester hypotesen $H: E(\bar{y}_{..}) = 0$, alternativt beskrivelse av tilpassningen av kun μ . Om tilpassningen forbedres med flere parametre enn kun μ er som før beskrevet ved en test med observatoren $F(R_m)$, der telleren er SSR_m som her er gitt ved:

$$(5.18) \quad R(\alpha, \beta, \gamma|\mu) = R(\mu, \alpha, \beta, \gamma) - R(\mu) = SSR_m = \sum_{i=1}^a \sum_{j=1}^b n_{ij} \bar{y}_{ij}^2 - n_{..} \bar{y}_{..}^2$$

Denne kan igjen dekomponeres for å finne ut vilke(n) faktor(er) som bidrar til eventuell signifikant forklaring av variasjonen i y . Da kan man enten starte med å tilpasse bare μ , deretter α etter μ ($R(\alpha|\mu)$), deretter β etter μ og α ($R(\beta|\mu, \alpha)$) og til slutt γ etter μ , α og β ($R(\gamma|\mu, \alpha, \beta)$), eller bytte om på rekkefølgen av α og β . Tilsvarende som i forrige avsnitt gjelder også dekomponeringen: $R(\alpha, \beta, \gamma|\mu) = R(\alpha|\mu) + R(\beta|\mu, \alpha) + R(\gamma|\mu, \alpha, \beta)$, der hvert av leddene er ikke-sentralt χ^2 -fordelt og representerer hver sine modeller som man kan teste effektiviteten av ved signifikantesting av de tilhørende F -observatorene. $R(\gamma|\mu, \alpha, \beta)$ er gitt fra (5.16) og (5.17) som:

$$(5.19) \quad R(\gamma|\mu, \alpha, \beta) = R(\mu, \alpha, \beta, \gamma) - R(\mu, \alpha, \beta)$$

og beskriver interaksjonsleddets bidrag til reduksjon i kvadratsummen utover de andre parametrene. Igjen som i forrige avsnitt har kvadratsummene også en tolkning som en testing av sammenhenger mellom parametrene. Den fremkommer etter å ha drøftet estimerbarhet nedenfor.

En opplagt begrensning til $R(\)$ notasjonen slik den er definert her, er noe som oppstår pga. at modellen ikke har full rang. Dette fremkommer når man har interaksjon med i i modellen og skyldes at hoved-effektene er "inneholdt" i interaksjonen (γ rommet inneholder både α og β rommet). Det gir seg utslag i $R(\)$ notasjonen på flg. måte:

$$(5.20) \quad R(\mu, \gamma) = R(\mu, \alpha, \gamma) = R(\mu, \beta, \gamma) = R(\mu, \alpha, \beta, \gamma)$$

Grunnen til dette skjer kan man se hvis man ser nøyere på f.eks. uttrykket $R(\mu, \alpha, \gamma)$ som tilsvarer en modell med interaksjonsledd, men uten den ene hoved-effekten. Denne modellen kan skrives: $y_{ijk} = \mu + \alpha_i + \gamma_{ij} + \varepsilon_{ijk}$, men dette er nettopp en 2-veis nestet modell. 2-veis nestede modeller har vi ikke tatt med her fordi resultatene for den lar seg lett utlede på samme måte som modellene vi har her, og inneholder ikke noe annet nytt. Imidlertid viser det seg at kvadratsumreduksjonen for en 2-veis nestet modell er (se [3], side 252):

$$(5.21) \quad R(\mu, \alpha, \gamma) = R(\mu, \beta, \gamma) = R(\mu, \gamma) = \sum_{i=1}^a \sum_{j=1}^b n_{ij} \bar{y}_{ij}^2$$

hvor de forskjellige variantene er μ , 1 hoved-effekt og interaksjons-ledd (nestet ledd), eller bare μ og interaksjons-ledd, som er et spesialtilfellet av de foregående, men som gir samme resultat fordi der en hoved-effekt er med blir denne likevel satt lik 0 i \mathbf{b}^0 og får dermed ikke bidrag i $SSR = \mathbf{b}' \mathbf{X}' \mathbf{y}$. Dette er altså alle modeller som ikke har begge hoved-effektene i tillegg til interaksjons-leddet. Vi ser umiddelbart at kvadratsumreduksjonen i en nestet modell er den samme som i vår 2-veis kryss-klassifisering med interaksjon, gitt i (5.16). Altså blir pr. definisjon:

$$(5.22) \quad \begin{aligned} R(\beta | \mu, \alpha, \gamma) &= R(\mu, \alpha, \beta, \gamma) - R(\mu, \alpha, \gamma) = 0 \\ R(\alpha | \mu, \beta, \gamma) &= R(\mu, \alpha, \beta, \gamma) - R(\mu, \beta, \gamma) = 0 \\ R(\alpha, \beta | \mu, \gamma) &= R(\mu, \alpha, \beta, \gamma) - R(\mu, \gamma) = 0 \end{aligned}$$

der disse $R(\)$ leddene på venstre side ikke er kvadratsummer i den betydningen vi har brukt hittil (siden de er 0). Dette betyr at bare når begge hoved-effektene er i modellen kan også interaksjons-leddet inkluderes, og først i det tilfellet får alle $R(\)$ leddene tolkning som meningsfulle kvadratsummer (og reduksjon i kvadratsummen).

Estimerbare funksjoner i denne modellen er som før basert på forventningene,

$E(y_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_{ij}$, så disse, eller lineær-kombinasjoner av de er estimerbare. For å forenkle notasjonen videre innfører vi $\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$ som betegnelse på forventningene. Først og fremst er μ_{ij} kun estimerbar i de rutene (kombinasjoner av i og j) det er data og estimatet ser vi fra (5.15) er gitt ved $\hat{\mu}_{ij} = \bar{y}_{ij}$, de eneste elementene i \mathbf{b}^0 som ikke er 0 er γ_{ij}^0 som er lik \bar{y}_{ij} . Der det mangler data er heller ikke μ_{ij} estimerbar. Individuelle parametre (μ, α_i, β_j) er ikke estimerbare. Heller ikke differanser mellom to α_i -er eller to β_j -er pga. at γ_{ij} vil ikke la seg eliminere. Hvis det er data i de respektive rutene er f.eks. $\mu_{11} - \mu_{21} = \alpha_1 - \alpha_2 + \gamma_{11} - \gamma_{21}$ estimerbar, men ikke bare $\alpha_1 - \alpha_2$.

Differanser mellom rad-effekter kan bare estimeres generelt med gjennomsnitt av kolonne-effekter og interaksjons-effekter til stede (kolonne-effektene forsvinner da hvis det er data i begge kolonnene).

Tilsvarende for kolonne-effekter. Noen av de generelle uttrykkene for estimerbare funksjonene er lange og lite intuitive, så derfor er de utelatt her (se [3], side 302). Det viktige er å huske at de kan alle formuleres som lineær-kombinasjoner av μ_{ij} .

Faktisk er det slik at ingen funksjon som kun inneholder α -er, β -er eller kombinasjoner av disse estimerbare. Derimot kan man finne estimerbare funksjoner som kun er avhengig av γ -er. Forutsatt at det er data i de respektive rutene, er:

$$(5.23) \quad \theta_{ij,i'j'} \equiv \mu_{ij} - \mu_{i'j} - (\mu_{ij'} - \mu_{i'j'}) = \gamma_{ij} - \gamma_{i'j} - (\gamma_{ij'} - \gamma_{i'j'})$$

estimerbar, og estimatet finnes ved å sette inn \bar{y}_{ij} med tilsvarende indekser. Noe som er interessant med (5.23) er at ikke bare funksjoner av estimerbare θ -er er estimerbare, men også noen funksjoner av ikke-estimerbare θ -er. F.eks. hvis data mangler i rute $(i, j) = (1, 2)$ er ikke μ_{12} estimerbar. Dermed er heller ikke $\theta_{11,22} = \mu_{11} - \mu_{21} - \mu_{12} + \mu_{22}$, eller $\theta_{12,33} = \mu_{12} - \mu_{32} - \mu_{13} + \mu_{33}$ estimerbar siden μ_{12} inngår i begge. Derimot er summen av disse estimerbar fordi μ_{12} forsvinner.

Uttrykkene i (5.16) og (5.17) gir oss alle mulige $R(\)$ som tilsvarende hver sin del-modell av vår 2-veis klassifisering med interaksjon. Signifikantesting av de forskjellige F -observatorene gir oss informasjon om hvilke faktorer som er viktige å ha med i modellen. Som i foregående avsnitt er de forskjellige $R(\)$ -leddene her også knyttet til konkrete hypoteser om parametrene i den fulle modellen. Disse hypotesene er mer omfattende enn tidligere, men nødvendige å ha med for sammenligning. De er utledet på samme måte som før, ved å ta utgangspunkt i den generelle test-observatoren fra (4.35) og vise i vilke tilfeller telleren i den reduseres til de forskjellige $R(\)$ -leddene.

Som i tidligere tilfeller er det også her tilstrekkelig å se på hypotesene fra (4.29) med $\mathbf{0}$ på høyre siden, (alle de relevante $R(\)$ leddene lar seg utlede av disse tilfellene). Den generelle hypotesen blir altså uttrykt som:

$$(5.24) \quad H: \mathbf{K}'\mathbf{b} = \mathbf{0}$$

Da er telleren i den generelle testobservatoren fra (4.36) $Q = (\mathbf{K}'\mathbf{b}^0)' (\mathbf{K}'\mathbf{G}\mathbf{K})^{-1} \mathbf{K}'\mathbf{b}^0$. Vi er interessert kun i de tilfellene med $\mathbf{K}'\mathbf{b}$ som estimerbare funksjoner og dermed lineære funksjoner av μ_{ij} -ene. Dermed kan vi skrive $\mathbf{K}'\mathbf{b} = \mathbf{L}'\boldsymbol{\mu}$ (med $\boldsymbol{\mu}$ som vektoren med μ_{ij} der $n_{ij} \neq 0$). Den delen av disse estimerbare funksjonene som inneholder γ_{ij} -ene vil nødvendigvis være på samme form som μ_{ij} -ene (altså $\mathbf{L}'\boldsymbol{\gamma}$), så det holder å se på hvordan γ_{ij} inngår i de estimerbare funksjonene for å lage \mathbf{L}' . Pga. den enkle formen på \mathbf{b}^0 og \mathbf{G} fra (5.15), der det kun er γ_{ij}^0 -ene som ikke er 0 (lik \bar{y}_{ij}) vil \mathbf{L}' kunne erstatte \mathbf{K}' i Q og ta vare på estimerbarheten. Med disse forenklingene er Q gitt ved:

$$(5.25) \quad Q = \bar{\mathbf{y}}'\mathbf{L} \left(\mathbf{L}'\mathbf{D} \begin{Bmatrix} \mathbf{1} \\ n_{ij} \end{Bmatrix} \mathbf{L} \right)^{-1} \mathbf{L}'\bar{\mathbf{y}}$$

der \mathbf{D} er diagonal-matrisen. Denne formen på Q lar seg så forenkle til $R(\)$ leddene vi kjenner igjen i forskjellige tilfeller.

Først ser vi på hypotesen som baseres på $F(M)$ som testobservator og husker fra før at den tester hypotesen $H: E(\bar{y}_{...}) = 0$. I vår modell er denne hypotesen uttrykt ved μ_{ij} lik

$$(5.26) \quad H: \sum_i \sum_j n_{ij} \mu_{ij} = 0 \text{ for } n_{ij} \neq 0$$

Siden $\mathbf{K}'\mathbf{b} = \mathbf{L}'\boldsymbol{\mu} = \mathbf{0}$ vil være formen på den generelle hypotesen vil $\mathbf{L}' = (n_{11} \ \dots \ n_{ab})$ formulere denne hypotesen. Da er det lett å se at Q i (5.25) reduseres til $R(\boldsymbol{\mu})$ i (5.17). Dette er bekreftelse på at $R(\boldsymbol{\mu})$ tester den samme hypotesen som før, og formulert ved parametrene.

Hypotesene som baseres på $F(\boldsymbol{\alpha}|\boldsymbol{\mu})$ (F -observatoren med $R(\boldsymbol{\alpha}|\boldsymbol{\mu})$ som teller) og $F(\boldsymbol{\beta}|\boldsymbol{\mu})$ (F -observatoren med $R(\boldsymbol{\beta}|\boldsymbol{\mu})$ som teller) er utledet på samme måte. Hypotesen gitt ved:

$$(5.27) \quad H: \alpha_i + \frac{1}{n_i} \sum_j n_{ij} (\beta_j + \gamma_{ij}) \text{ like for alle } i$$

kan lett vises at testes ved $R(\boldsymbol{\alpha}|\boldsymbol{\mu})$. Først skrives hypotesen som de $a-1$ differanser

$\frac{1}{n_i} \sum_j n_{1j} \mu_{1j} - \frac{1}{n_i} \sum_j n_{ij} \mu_{ij} = 0, \quad i = 2, 3, \dots, a$. Så konstrueres \mathbf{L}' ut fra denne formuleringen og etter litt regning ([3], s 307) ser man at Q i (5.25) reduseres til $R(\boldsymbol{\alpha}|\boldsymbol{\mu})$ (utregnet fra 5.17). Det vil si at $F(\boldsymbol{\alpha}|\boldsymbol{\mu})$ tester hypotesen i (5.27). Tilsvarende er det for $F(\boldsymbol{\beta}|\boldsymbol{\mu})$. Denne tester hypotesen gitt ved:

$$(5.28) \quad H: \beta_j + \frac{1}{n_j} \sum_i n_{ij} (\alpha_i + \gamma_{ij}) \text{ like for alle } j$$

Det er verdt å legge merke til at hvis $\gamma_{ij} = 0$ i (5.27) og (5.28) får vi samme hypotesene som i forrige avsnitt (drøftet helt på slutten), noe som er forventet siden modellen der var uten interaksjon.

Hypotesen som testes ved $F(\boldsymbol{\alpha}|\boldsymbol{\mu}, \boldsymbol{\beta})$ ($R(\boldsymbol{\alpha}|\boldsymbol{\mu}, \boldsymbol{\beta})$) er gitt ved:

$$(5.29) \quad H: \left(n_i - \sum_{j=1}^b \frac{n_{ij}^2}{n_j} \right) \alpha_i - \sum_{i' \neq i}^a \left(\sum_{j=1}^b \frac{n_{ij} n_{i'j}}{n_j} \right) \alpha_{i'} + \sum_{j=1}^b \left(n_{ij} - \frac{n_{ij}^2}{n_j} \right) \gamma_{ij} - \sum_{i' \neq i}^a \sum_{j=1}^b \left(\frac{n_{ij} n_{i'j}}{n_j} \right) \gamma_{i'j} = 0, \quad i = 1, \dots, a-1$$

her går i bare til $a-1$ fordi summen av venstresiden over alle i er 0, så for at hypotesene skal være uavhengige brukes bare de $a-1$ første.

Helt tilsvarende gjelder for $F(\boldsymbol{\beta}|\boldsymbol{\mu}, \boldsymbol{\alpha})$ ($R(\boldsymbol{\beta}|\boldsymbol{\mu}, \boldsymbol{\alpha})$) som tester hypotesen:

$$(5.30) \quad H: \left(n_j - \sum_{i=1}^a \frac{n_{ij}^2}{n_i} \right) \beta_j - \sum_{j' \neq j}^b \left(\sum_{i=1}^a \frac{n_{ij} n_{ij'}}{n_i} \right) \beta_{j'} + \sum_{i=1}^a \left(n_{ij} - \frac{n_{ij}^2}{n_i} \right) \gamma_{ij} - \sum_{j' \neq j}^b \sum_{i=1}^a \left(\frac{n_{ij} n_{ij'}}{n_i} \right) \gamma_{ij'} = 0, \quad j = 1, \dots, b-1$$

Disse utledes på samme måte som over ved først å uttrykke hypotesene på formen $\mathbf{L}'\boldsymbol{\mu} = \mathbf{0}$. Deretter kan matrisen \mathbf{L}' konstrueres på grunnlag av koeffisientene til γ_{ij} og ved utregning ses at Q i (5.25) reduseres til $R(\alpha|\mu, \beta)$ i (5.29) og til $R(\beta|\mu, \alpha)$ i (5.30), når $R(\alpha|\mu, \beta)$ og $R(\beta|\mu, \alpha)$ regnes ut ved uttrykkene i (5.17). Tilsvarende som ovenfor fås at når vi setter $\gamma_{ij} = 0$ f.eks. i (5.30) fremkommer en ligning som bare kan være oppfylt når alle β -ene er like, noe som tilsvarer den hypotesen som $F(\beta|\mu, \alpha)$ tester i forrige avsnitt (der modellen var uten interaksjon).

Til slutt er hypotesen som testes ved $F(\gamma|\mu, \alpha, \beta)$ ($R(\gamma|\mu, \alpha, \beta)$) gitt ved:

$$(5.31) \quad H: \left. \begin{array}{l} \text{En vilkårlig kolonnevektor med } s - a - b + 1 \text{ lineært uavhengige} \\ \text{funksjoner av } \theta_{ij,i'j'} \text{ der disse funksjonene er enten estimerbare} \\ \theta\text{-er, eller estimerbare summer / differanser av } \theta\text{-er} \end{array} \right\} = \mathbf{0}$$

der s er antall ruter det er data i, $s - a - b + 1$ er antall frihetsgrader i $R(\gamma|\mu, \alpha, \beta)$, altså hvor mange uavhengige ledd det er i kvadratsummen, og $\theta_{ij,i'j'}$ er definert i (5.23). For å skissere hvordan det vises at hypotesen i (5.31) testes med $F(\gamma|\mu, \alpha, \beta)$ ($R(\gamma|\mu, \alpha, \beta)$) omformuleres først hypotesen (siden det bare er γ -ledd som inngår) til formen: $\mathbf{L}'\boldsymbol{\gamma} = \mathbf{0}$, der \mathbf{L}' har dimensjon $(s - a - b + 1) \times s$ og full rad-rang. Ved å se på definisjonen av θ i (5.23) ser vi at betraktet som et ligningssystem er en løsning til dette gitt ved kolonnevektoren av bare 1-ere ($\mathbf{1}$), slik at $\mathbf{L}'\mathbf{1} = \mathbf{0}$. Da må også $\boldsymbol{\gamma}\mathbf{1}$ være en løsning, altså $\gamma_{ij} = \gamma$ for alle i og j der $n_{ij} \neq 0$. Men med konstant interaksjonsledd kan modellen under hypotesen i (5.31) betraktes som en redusert modell: $E(y_{ijk}) = (\mu + \gamma) + \alpha_i + \beta_j$ med reduksjon av kvadratsummen lik $R(\mu, \alpha, \beta)$. Når hypotesen er det eneste som skiller full og redusert modell kan da Q som er teller-kvadratsummen til F -observatoren for den generelle hypotesen fra (4.35) betraktes som endringen i kvadratsumreduksjon som skyldes hypotesen, altså $Q = R(\mu, \alpha, \beta, \gamma) - R(\mu, \alpha, \beta)$ (se [3], side 119), men dette er jo nettopp lik $R(\gamma|\mu, \alpha, \beta)$. Dermed viser dette at $F(\gamma|\mu, \alpha, \beta)$ tester hypotesen i (5.31).

Et viktig moment, som kun gjelder i det ubalanserte tilfellet (forskjellig antall observasjoner i rutene) er knyttet til tolkningen av $F(\gamma|\mu, \alpha, \beta)$. En test der $F(\gamma|\mu, \alpha, \beta)$ ikke er signifikant, betyr ikke at det ikke er interaksjon i modellen. Grunnen til dette er lettest å se fra definisjonen av θ i (5.23) sammen med hypotesen (5.31). θ ser vi kan også skrives

$\theta_{ij,i'j'} = E(\bar{y}_{ij.}) - E(\bar{y}_{i'.j.}) - [E(\bar{y}_{ij'.}) - E(\bar{y}_{i'j'.})]$ som beskriver i hvilken grad forskjellen i forventet rad-effekt mellom i og i' varierer mellom kolonne j og j' , som nettopp er definisjonen av interaksjon.

Dersom hypotesen i (5.31) hadde vært å teste $\theta_{ij,i'j'} = 0$ for forskjellige i, j, i', j' ville dette gitt svar på om dataene tydet på signifikant interaksjon, eller ikke. Men hypotesen tillater også at f.eks. summen av to $\theta_{ij,i'j'}$ kan være 0, slik vi så i eksempelet etter (5.23), hvor summen av to $\theta_{ij,i'j'}$ kan være estimerbar, selv om de ikke hver for seg er det, siden en μ_{ij} som ikke er estimerbar lar seg eliminere. Det er faktisk mønsteret av tomme celler (de cellene som mangler data) som avgjør hvilke funksjoner av γ_{ij} som er estimerbare og som kan brukes i hypotesen (5.31). Hvis summen av to $\theta_{ij,i'j'}$ er 0 trenger jo ikke hver av dem være 0, de kan være like store med motsatt fortegn, noe som gjør at det kan godt være interaksjoner i modellen uten at $F(\gamma|\mu, \alpha, \beta)$ er signifikant. For å finne ut av dette må en

spesiell hypotese spesifiseres for $\theta_{ij,i'j'} = 0$ og testes ved (5.24) og (5.25). Hvis $F(\gamma|\mu, \alpha, \beta)$ er signifikant tyder det derimot alltid på at det er interaksjon forskjellig fra 0.

6. Oppklaring - hårfin balanse

I avsnitt 2.3 ble forskjellen på ubalanserte og balanserte data beskrevet med et eksempel. I det balanserte tilfellet er det like mange observasjoner i hver gruppe på finest oppdelte nivå (rute, celle, eller kombinasjon av i og j), mens det i det ubalanserte tilfellet ikke er det. Etter å ha gått gjennom de forskjellige kvadratsumreduksjonene man kan splitte en 2-veis kryss klassifisering med interaksjon i , i kapittel 5, kan vi nå forklare hvor forskjellen mellom balanserte og ubalanserte data spiller en rolle i statistisk inferens.

Egentlig er det ikke nødvendig å skille dem, man kan betrakte det balanserte tilfellet som et spesialtilfelle av det ubalanserte, som er det Searle gjør i [3] og det mest logiske siden formlene for det balanserte tilfelle lar seg utlede fra det ubalanserte og ikke omvendt. Likevel har det historisk vært det balanserte tilfellet som først ble grundig analysert, og det har lenge vært enighet om hvilken variansanalyse som er den riktige. De relevante hypotesene å teste er hoved-effekter og interaksjon (se [6]). Det ubalanserte tilfellet har så vært betraktet som en utvidelse av det balanserte, men det har ikke vært den samme "konsensus" for analyse / hypoteser, og så seint som på 1970-tallet var det uenighet (se [5,6]). Pga. forenklingene i formlene for det balanserte tilfellet, er det f.eks. i SAS laget prosedyrer som bare er beregnet for balanserte data, som er beregningsmessig raskere enn de mer generelle prosedyrene, og også lettere å forstå. Enkelheten er et viktig moment både for lett å kunne forstå riktig måte å lese utskriftene på, men også en hjelp når anvendelsene skal illustreres. I de mer generelle prosedyrene for ubalanserte data beregnes fire forskjellige kvadratsummer (som for øvrig blir like for balanserte data) som skal dekke alle varianter av regresjons- og variansanalyse. Å forstå forskjellen på disse fire kvadratsummene er nødvendig for å analysere ubalanserte data i SAS. I tillegg til at det balanserte tilfellet var forstått først og at programpakken senere bevisst har opprettholdt forskjellen på balansert / ubalansert tilfelle, er det også slik at i mange lærebøker gis det balanserte tilfellet mye oppmerksomhet for lett å kunne forstå grunnprinsippene. Det er derfor nyttig å repetere hvordan noen av formlene i det balanserte tilfellet ser ut, for deretter og peke på forenklingene i forhold til det ubalanserte tilfellet.

6.1 Balanserte data

Vi tenker oss igjen en 2-veis kryssklassifisering med interaksjon, gitt ved $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$, der som før $i=1, \dots, a$, $j=1, \dots, b$ og $k=1, \dots, n$ slik at n er antall observasjoner i hver rute. Videre trenger vi ligningen (fra grunnkurs), som også gjelder for ubalanserte data:

$$(6.1) \quad \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (y_{ijk} - \bar{y}_{...})^2 = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n y_{ijk}^2 - \frac{1}{abn} \left(\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n y_{ijk} \right)^2 = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n y_{ijk}^2 - abn \bar{y}_{...}^2$$

Så skriver vi om:

$$(6.2) \quad (y_{ijk} - \bar{y}_{...})^2 = \left[(\bar{y}_{i..} - \bar{y}_{...}) + (\bar{y}_{.j.} - \bar{y}_{...}) + (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}) + (y_{ijk} - \bar{y}_{ij.}) \right]^2$$

som antyder en dekomposisjon i de 4 leddene på høyre-siden. Det som gjør utregningene enkle i det balanserte tilfellet er at når (6.2) summeres over i, j, k splittes kvadratsummen kun opp i de 4 leddene i paranteser på høyresiden, slik at vi får:

$$\begin{aligned}
(6.3) \quad \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (y_{ijk} - \bar{y}_{...})^2 &= \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{y}_{i..} - \bar{y}_{...})^2 + \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{y}_{.j.} - \bar{y}_{...})^2 \\
&+ \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 + \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (y_{ijk} - \bar{y}_{ij.})^2 \\
&= bn \sum_{i=1}^a (\bar{y}_{i..} - \bar{y}_{...})^2 + an \sum_{j=1}^b (\bar{y}_{.j.} - \bar{y}_{...})^2 \\
&+ n \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 + \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (y_{ijk} - \bar{y}_{ij.})^2
\end{aligned}$$

Alle kryss-produktene mellom de 4 leddene faller altså bort (blir 0), men bare i det balanserte tilfellet. Grunnen til dette er at med $n_{ij} = n$ (balansert tilfelle) er f.eks. $\sum_i \bar{y}_{...} \sum_j n_{ij} \bar{y}_{.j.} = \sum_i \bar{y}_{i..} \sum_j n_{ij} \bar{y}_{.j.}$, og ikke ellers. I (6.1) kjenner vi igjen (på høyresiden) uttrykkene som den totale kvadratsummen og $R(\mu)$. Før vi skal se på sammenhengen med formlene i kapittel 5 skal vi vise at i det balanserte tilfellet er faktisk

$$\begin{aligned}
(6.4) \quad (i) \quad R(\beta | \mu, \alpha) &= R(\beta | \mu) \\
(ii) \quad R(\alpha | \mu, \beta) &= R(\alpha | \mu)
\end{aligned}$$

Vi ser bare på den øverste av disse og viser det bare for $n = 1$ (for enkelhets skyld). Da forsvinner interaksjonsleddet, noe som er en viktig observasjon i seg selv, det at interaksjon slik vi har definert det i kapittel 5, kun gir mening i de tilfellene man har flere observasjoner i hver rute. Dette er siden det ikke er mulig å skille mellom γ_{ij} og ε_{ijk} når k er overflødig, og da er det nok med en av disse som blir støyleddet i den nye modellen. Modellen er altså $y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$ som i avsnitt 5.3. $R(\beta | \mu, \alpha)$ er gitt i (5.14) og vi trenger den inverse \mathbf{C} matrisen. Fra (5.9) har vi $c_{jj} = a - \frac{a}{b}$, $c_{jj'} = -\frac{a}{b}$ som gir:

$$(6.5) \quad \mathbf{C} = a\mathbf{I} - \frac{a}{b}\mathbf{J}$$

der \mathbf{J} er en matrise med bare 1-ere. Da kan et resultat fra lineæralgebraen brukes for å finne den inverse (se f.eks. [3] side 327), nemlig at:

$$(k_1 \mathbf{I}_m + k_2 \mathbf{J}_m)^{-1} = \frac{1}{k_1} \left(\mathbf{I}_m - \frac{k_2}{k_1 + mk_2} \mathbf{J}_m \right)$$

Innsatt her får vi (her er dimensjonen $b - 1$):

$$\mathbf{C}^{-1} = \frac{1}{a} (\mathbf{I} + \mathbf{J}) = \begin{pmatrix} \frac{2}{a} & \frac{1}{a} & \dots & \frac{1}{a} \\ \frac{1}{a} & \ddots & & \vdots \\ \vdots & & \ddots & \frac{1}{a} \\ \frac{1}{a} & \dots & \frac{1}{a} & \frac{2}{a} \end{pmatrix}$$

Fra (5.10) har vi at $r_j = \left\{ a(\bar{y}_{.j.} - \bar{y}_{...}) \right\}$, for $j = 1, \dots, b - 1$. Da fås

$$\boldsymbol{\beta}^0 = \mathbf{C}^{-1} \mathbf{r} = \begin{pmatrix} \frac{2}{a} & \frac{1}{a} & \cdots & \frac{1}{a} \\ \frac{1}{a} & \ddots & & \vdots \\ \vdots & & \ddots & \frac{1}{a} \\ \frac{1}{a} & \cdots & \frac{1}{a} & \frac{2}{a} \end{pmatrix} a \begin{pmatrix} \bar{y}_1 - \bar{y}_.. \\ \bar{y}_2 - \bar{y}_.. \\ \vdots \\ \bar{y}_{(b-1)} - \bar{y}_.. \end{pmatrix} = \begin{pmatrix} 2\bar{y}_1 + \bar{y}_2 + \cdots + \bar{y}_{b-1} - b\bar{y}_.. \\ \bar{y}_1 + 2\bar{y}_2 + \cdots + \bar{y}_{b-1} - b\bar{y}_.. \\ \vdots \\ \bar{y}_1 + \bar{y}_2 + \cdots + 2\bar{y}_{b-1} - b\bar{y}_.. \end{pmatrix} = \begin{pmatrix} \bar{y}_1 - \bar{y}_b \\ \bar{y}_2 - \bar{y}_b \\ \vdots \\ \bar{y}_{(b-1)} - \bar{y}_b \end{pmatrix}$$

hvor vi har brukt at $b\bar{y}_.. = \bar{y}_1 + \cdots + \bar{y}_b$. Til slutt fås:

$$R(\boldsymbol{\beta}|\boldsymbol{\mu}, \boldsymbol{\alpha}) = \boldsymbol{\beta}^{0'} \mathbf{r} = ((\bar{y}_1 - \bar{y}_b) \quad \cdots \quad (\bar{y}_{(b-1)} - \bar{y}_b)) a \begin{pmatrix} (\bar{y}_1 - \bar{y}_..) \\ \vdots \\ (\bar{y}_{(b-1)} - \bar{y}_..) \end{pmatrix}$$

Dette produktet gir $(b-1)$ ledd av typen $a(\bar{y}_j - \bar{y}_b)(\bar{y}_j - \bar{y}_..) = a[\bar{y}_j^2 - \bar{y}_j\bar{y}_.. - \bar{y}_b\bar{y}_j + \bar{y}_b\bar{y}_..]$ for $j=1, \dots, b-1$. Det første leddet i klammeparentesen summeres opp til $\sum_{j=1}^{b-1} \bar{y}_j^2$. Det andre leddet sammen med det fjerde summeres til: $-\bar{y}_..(\bar{y}_1 + \cdots + \bar{y}_{b-1} - (b-1)\bar{y}_b) = -b\bar{y}_..^2 + b\bar{y}_b\bar{y}_..$ siden $b\bar{y}_.. = \bar{y}_1 + \cdots + \bar{y}_b$ som over. Det tredje leddet summeres til $-\bar{y}_b(\bar{y}_1 + \cdots + \bar{y}_{b-1}) = -\bar{y}_b(b\bar{y}_.. - \bar{y}_b) = -b\bar{y}_b\bar{y}_.. + \bar{y}_b^2$. Til sammen blir resultatet at \bar{y}_b^2 går inn i summen, to ledd faller mot hverandre og vi står igjen med uttrykket $a(\sum_{j=1}^b \bar{y}_j^2 - b\bar{y}_..^2)$ som nettopp er uttrykket for $R(\boldsymbol{\beta}|\boldsymbol{\mu})$ i (5.7) når det er en $\boldsymbol{\beta}$ -faktor i modellen med $n_j = a$ observasjoner på hvert nivå. Dermed har vi vist (i) i (6.4).

Tolkningen av (6.4) fortjener oppmerksomhet og er den viktigste forskjellen mellom balansert og ubalansert tilfellet. Som vi husker er $R(\boldsymbol{\beta}|\boldsymbol{\mu}, \boldsymbol{\alpha})$ hvor mye mer kvadratsummen minker når man tilpasser $\boldsymbol{\beta}$ i en modell som allerede har $\boldsymbol{\alpha}$ og $\boldsymbol{\mu}$. At denne er det samme som $R(\boldsymbol{\beta}|\boldsymbol{\mu})$ betyr at det spiller ingen rolle om man har $\boldsymbol{\alpha}$ i modellen eller ikke når man tilpasser $\boldsymbol{\beta}$ (og tilsvarende i motsatt rekkefølge), noe vi vet ikke er tilfelle med ubalanserte data. Konsekvensen av dette er at ANOVA tabellen blir enklere, for man trenger bare betrakte $R(\boldsymbol{\alpha}|\boldsymbol{\mu})$ og $R(\boldsymbol{\beta}|\boldsymbol{\mu})$ i motsetning til det ubalanserte tilfellet der man først tilpasser den ene effekten etter $\boldsymbol{\mu}$ f.eks. $R(\boldsymbol{\alpha}|\boldsymbol{\mu})$, og deretter tilpasser $\boldsymbol{\beta}$ etter $\boldsymbol{\mu}$ og $\boldsymbol{\alpha}$ ved $R(\boldsymbol{\beta}|\boldsymbol{\mu}, \boldsymbol{\alpha})$, og så til slutt i motsatt rekkefølge, altså $R(\boldsymbol{\beta}|\boldsymbol{\mu})$ og deretter $R(\boldsymbol{\alpha}|\boldsymbol{\mu}, \boldsymbol{\beta})$ for å dekke alle mulighetene. Litt upresist kan man også si at $R(\boldsymbol{\beta}|\boldsymbol{\mu}, \boldsymbol{\alpha})$ er bidraget fra faktoren $\boldsymbol{\beta}$, etter å ha justert for $\boldsymbol{\mu}$ og $\boldsymbol{\alpha}$. Da betyr (6.4) (i) at siden det ikke spiller noen rolle om man justerer for $\boldsymbol{\alpha}$ eller ikke, vil $\boldsymbol{\alpha}$ effekten ikke influere på $\boldsymbol{\beta}$ effekten i det hele tatt og motsatt. Med ubalanserte data er det som sagt ikke slik, og man kan si at effektene blander seg, noe eksempelet i neste avsnitt er en god illustrasjon på.

Nå er vi i stand til å sette opp den relevante variansanalysen (ANOVA tabellen) i det balanserte tilfellet og koble den til de mer generelle formlene fra kapittel 5. Vi begynner med (6.1) som altså er uttrykk for den totale variasjonen i dataene. Om vi bruker avvikene fra gjennomsnittet, eller den totale kvadratsummen er ikke så farlig, men siden vi er vant til å bruke den totale kvadratsummen til dataene i kapittel 5, gjør vi det her også. Da kjenner vi igjen uttrykkene på høyresiden i (6.1) som

$$(6.6) \quad SST = \sum_i \sum_j \sum_k y_{ijk}^2 \quad \text{og} \quad R(\boldsymbol{\mu}) = abn \bar{y}_..^2$$

Videre har vi med notasjonen fra kapittel 5 at for modellen med interaksjon gjelder at:

$R(\mu, \alpha, \beta, \gamma) = R(\gamma|\mu, \alpha, \beta) + R(\beta|\mu, \alpha) + R(\alpha|\mu) + R(\mu)$ og med (6.4) får vi den unike dekomponeringen i det balanserte tilfellet:

$$(6.7) \quad R(\mu, \alpha, \beta, \gamma) = R(\gamma|\mu, \alpha, \beta) + R(\beta|\mu) + R(\alpha|\mu) + R(\mu)$$

Ved samme resonnement som i omformingen i (6.1) har vi at

$R(\beta|\mu) = a \sum_{j=1}^b \bar{y}_j^2 - ab \bar{y}_{..}^2 = \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{.j} - \bar{y}_{..})^2$ som også blir på samme form med n observasjoner i stedet for 1 observasjon i hver rute, slik at vi får

$$(6.8) \quad R(\beta|\mu) = \sum_i \sum_j \sum_k (\bar{y}_{.j} - \bar{y}_{..})^2 = an \sum_j (\bar{y}_{.j} - \bar{y}_{..})^2$$

og tilsvarende:

$$(6.9) \quad R(\alpha|\mu) = \sum_i \sum_j \sum_k (\bar{y}_{i..} - \bar{y}_{..})^2 = bn \sum_i (\bar{y}_{i..} - \bar{y}_{..})^2$$

Disse uttrykkene kjenner vi igjen fra (6.3). At effektene ikke blander seg, kan man få et intuitivt inntrykk av ved å se på f.eks. $R(\beta|\mu)$ som bare inneholder kolonne-gj.nitt (ikke rad-gj.snitt). På høyre siden av (6.3) har vi også uttrykket $\sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{ij.})^2$ som vi ser beskriver variasjonen i dataene innen rutene (minste nivå), og altså det som modellstrukturen ikke fanger opp (rader, kolonner og interaksjoner mellom disse). Derfor er det ikke overraskende at dette leddet nettopp er *SSE*.

$$(6.10) \quad SSE = \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{ij.})^2$$

Hvis *SSE* på høyresiden av (6.3) bytter plass med $R(\mu)$ fra (6.6) får vi på venstre siden av (6.3) $SST - SSE$ som vi kjenner igjen som definisjonen på *SSR*, eller $R(\mu, \alpha, \beta, \gamma)$. Det som gjenstår å gjøre rede for på høyresiden av (6.3) må da fra (6.7) være $R(\gamma|\mu, \alpha, \beta)$, altså:

$$(6.11) \quad R(\gamma|\mu, \alpha, \beta) = n \sum_i \sum_j (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j} + \bar{y}_{..})^2$$

Ligningene (6.6), (6.8), (6.9), (6.10) og (6.11) utgjør variansanalysen (ANOVA tabellen) for balanserte data. (6.3) inneholder det samme, og vi får det på formen (6.7) ved å splitte opp venstresiden som i (6.1) og bytte $R(\mu)$ over på høyresiden og *SSE* over på venstre, ev. ved bare å la *SST* stå alene på venstresiden illustreres dekomposisjonen som i en ANOVA tabell.

6.2 Eksempel

Et enkelt eksempel er egnet for å illustrere noe av det som ble beskrevet i forrige avsnitt om konsekvensene av at dataene ikke er balansert. Vi tenker oss dataene som i tabellen nedenfor:

		B	
		1	2
A	1	7, 9	5
	2	8	4, 6

Tabell 7: Ubalanserte data

Vi har faktorene A (radene) og B (kolonnene) med 2 nivåer hver. Hvis vi ser på nivå 1 av B er det ingenting som tyder på forskjell mellom rad 1 og 2 i A siden $\bar{y}_{11} = (7+9)/2 = 8$ og $\bar{y}_{21} = 8$.

Tilsvarende er det ingenting som tyder på radforskjell for nivå 2 av B siden $\bar{y}_{12} = \bar{y}_{22} = 5$. Det er altså ikke forskjell på radene for noe nivå av B . Hvis man derimot ser på radgjennomsnittene er disse $\bar{y}_{1..} = (7+9+5)/3 = 7$ og $\bar{y}_{2..} = (8+4+6)/3 = 6$ så forskjellen på disse kan tolkes som en forskjell på nivåene i A . Men i virkeligheten måler radgjennomsnittene både effekten av faktor B og faktor A .

Dette ser vi hvis vi setter skriver modellen som: $y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}$ (dropper interaksjonsledd for enkelhets skyld). Da kan vi sette opp radgjennomsnittene som

$$\begin{aligned}\bar{y}_{1..} &= \left(\frac{1}{3}\right)[3\mu + 2(\alpha_1 + \beta_1) + (\alpha_1 + \beta_2)] = \mu + \alpha_1 + \frac{2}{3}\beta_1 + \frac{1}{3}\beta_2 \\ \bar{y}_{2..} &= \left(\frac{1}{3}\right)[3\mu + (\alpha_2 + \beta_1) + 2(\alpha_2 + \beta_2)] = \mu + \alpha_2 + \frac{1}{3}\beta_1 + \frac{2}{3}\beta_2\end{aligned}$$

og da blir differansen mellom disse

$$(6.12) \quad \bar{y}_{1..} - \bar{y}_{2..} = (\alpha_1 - \alpha_2) + \frac{1}{3}(\beta_1 - \beta_2)$$

som viser at differansen av radgjennomsnittene ikke estimerer differansen mellom nivåene i faktor A , $(\alpha_1 - \alpha_2)$, men dette + en skalert differanse mellom nivåene i B .

Hvis vi heller formulerer modellen som $y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$, der man vanligvis i denne 2×2 tabellen ville estimert nivåene i A som $(\mu_{11} + \mu_{12})/2$ og $(\mu_{21} + \mu_{22})/2$, og differansen mellom nivåene som differansen mellom disse, blir radgjennomsnittene gitt ved

$$\begin{aligned}\bar{y}_{1..} &= (2\mu_{11} + \mu_{12})/3 + \bar{\varepsilon}_{1..} \\ \bar{y}_{2..} &= (\mu_{21} + 2\mu_{22})/3 + \bar{\varepsilon}_{2..}\end{aligned}$$

som på sin side estimerer $(2\mu_{11} + \mu_{12})/3$ og $(\mu_{21} + 2\mu_{22})/3$. Disse er funksjoner av antall observasjoner i rutene, noe som kan være uheldig, f.eks. i hypotesetesting hvor man helst skal kunne sette opp den interessante hypotesen før man vet hvor mange observasjoner som havner i de forskjellige kategoriene.

Hvis vi i stedet ser på de balanserte dataene i tabellen nedenfor får vi for nivå 1 i B :

$\bar{y}_{11} = (7+9)/2 = 8$ og $\bar{y}_{21} = (6+10)/2 = 8$ og for nivå 2 i B : $\bar{y}_{12} = \bar{y}_{22} = 5$ som i det ubalanserte tilfellet. Nå blir radgjennomsnittene $\bar{y}_{1..} = (7+9+3+7)/4 = 13/2$ og $\bar{y}_{2..} = (6+10+4+6)/4 = 13/2$ og siden disse er like kan vi tolke det som at det ikke er forskjell på nivåene i A .

		B	
		1	2
A	1	7, 9	3, 7
	2	6, 10	4, 6

Tabell 8: Balanserte data

Setter vi inn i modellen, får vi at

$$\bar{y}_{1..} = \left(\frac{1}{4}\right)[4\mu + 2(\alpha_1 + \beta_1) + 2(\alpha_1 + \beta_2)] = \mu + \alpha_1 + \frac{1}{2}\beta_1 + \frac{1}{2}\beta_2$$

$$\bar{y}_{2..} = \left(\frac{1}{4}\right)[4\mu + 2(\alpha_2 + \beta_1) + 2(\alpha_2 + \beta_2)] = \mu + \alpha_2 + \frac{1}{2}\beta_1 + \frac{1}{2}\beta_2$$

som gir at differansen mellom disse blir

$$\bar{y}_{1..} - \bar{y}_{2..} = \alpha_1 - \alpha_2$$

som er det vi vil estimere og måler forskjellen i nivå for faktor A uten innblanding fra faktor B .

Hovedproblemet i analyse av ubalanserte data er at en differanse mellom gjennomsnitt i en faktor blir influert av effekten av en annen faktor, og løsningen på dette er å korrigere gjennomsnittene for dette.

7. Mer oppklaring - lunefull entydighet

I avsnitt 2.3 så vi på et eksempel der det ikke fantes entydig løsning av normallikningene. Når man har flere parametere enn kategorier i dataene å estimere dem fra er det alltid slik, modellen har ikke full rang. Modellen uten full rang er likevel vanlig å bruke pga. enkel tolkning av parametrene. Det har derfor vært viktig å finne ut hvordan denne overparametriseringen påvirker inferensen (se [7]). En annen måte å angripe mangelen på entydighet på, er å lage restriksjoner (tilleggslikninger) på parametrene slik at man får entydighet (se [7]). En tredje måte er å unngå problemet ved å bruke andre modeller som har full rang (se [6], [8]). Uansett overparametrisering eller ikke baserer variansanalysen seg på en dekomponering av kvadratsummer. Det er som vi så i kapittel 5, en nær forbindelse mellom kvadratsumreduksjoner og hypoteser om parametre i modellen. Vi skal se her at avhengig av hvordan man løser problemet med overparametrisering, har valgene store konsekvenser for kvadratsummene og dermed for hvilke hypoteser som lar seg teste, noe som ikke nødvendigvis går frem av $R(\)$ notasjonen som vi har brukt. Man må derfor vite om de forskjellige metodene og forstå forskjeller og likheter. Forvirringspotensialet er stort, delvis pga. overparametriseringen, delvis pga. at hypotesene, ikke er entydig spesifisert ved $R(\)$ notasjonen, og delvis pga. at det kan være vanskelig å tolke hypotesene. Ved å beskrive de tre forskjellige metodene og de tilhørende hypotesene kan dette forhåpentlig bidra til å dekomponere forvirringen i mindre biter som hver for seg er lettere å bryte ned.

Vi har sett at ved hjelp av lineær algebra og den generaliserte inversen, er ikke mangel på entydighet noe problem rent regneteknisk. Den manglende entydigheten får imidlertid konsekvenser (begrensninger) for estimeringen, men ved å studere de såkalte estimerbare funksjonene finner man ut både hva som kan estimeres og hvilke hypoteser som lar seg teste under disse modellene. Denne første

metoden "drar" altså med seg overparametriseringen. Det er flere fordeler med dette. Den generaliserte inversen lar seg alltid regne ut og de relevante kvadratsumreduksjonene, de forskjellige $R(\cdot)$ -leddene, er alltid definert og unike. Likevel har de sine begrensninger som vi skal se nedenfor. Vi kaller denne metoden for **Metode 1**.

Å lage "kunstige" restriksjoner på parametrene i modellen er også en enkel fremgangsmåte for å oppnå entydig løsning av normalligningene. Denne metoden har opplagte ulemper ved at restriksjonene påfører modellen begrensninger som må tas med når man tolker parametrene og resultatene, restriksjoner som kanskje ikke passer de aktuelle dataene man har. Med modellen i (4.3), 2-veis kryssklassifisering med interaksjon er et eksempel på en restriksjon $\alpha_i = \beta_j = \gamma_{ij} = 0$ for alle i, j . Da får modellen full rang og normalligningene har entydig løsning. Dette kaller vi **Metode 2**. Hva i all verden er det så som skiller denne metoden fra Metode 1 ovenfor? Vi husker at i Metode 1 for å regne ut den generaliserte inversen (avsnitt 5.4) var det nettopp slik vi gjorde det, vi valgte noen parametre lik 0 og løste et ligningssystem med full rang og fikk løsningen i (5.15). Men det er faktisk stor forskjell på disse metodene, de har forskjellig tolkning og de gir forskjellige resultater. Vi skal se nedenfor at kvadratsumreduksjonene i Metode 2 ikke alltid er unike, to forskjellige sett med restriksjoner er opphav til forskjellige hypoteser.

Den siste metoden vi skal se på, kaller vi **Metode 3** og er å i stedet for å analysere modellen fra avsnitt 5.4 uten full rang, heller betrakte modellen $y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$. Denne har full rang i tillegg til at parametrene er lette å tolke, de er nemlig populasjonsgjennomsnittene. Vi betraktet denne modellen i avsnitt 5.4 da vi av notasjonsmessige hensyn satte $\mu_{ij} = \mu + \alpha_i + \beta_j$, men da fokuserte vi ikke på at dette samtidig representerte en modell med full rang. De andre modellene vi har sett på lar seg også representere ved denne modellen, med passende restriksjoner. Vi presenterer disse restriksjonene slik at forbindelsen mellom denne og de andre modellene er klare, og beskriver hvordan hypotesene under denne modellen ser ut, ettersom noen av hypotesene er lettere å tolke med disse parametrene (som i [6]). Kvadratsumreduksjoner i disse modellene lar seg regne ut på tilsvarende måte som vi har gjort i de overparametriserte modellene våre (se [8]), men dette bidrar ikke med noe nytt, utover at man blir bedre kjent med denne type modeller. Vi har først og fremst sett nytten av å formulere hypotesene under disse modellene siden tolkningen da kan være mer intuitiv, og nøyer oss utover dette å beskrive forbindelsen med de overparametriserte modellene. Denne typen modeller ble mer vanlig først på 1980 tallet, etter at forvirringen knyttet til de overparametriserte modellene hadde skapt debatt i 10-året før. Det er viktig å huske at både med og uten restriksjoner har denne modellen full rang.

Generelt

Først repeterer vi notasjonen og antagelsene som er felles for Metode 1 og Metode 2 som i [7] for å beskrive og forklare forskjellene mellom disse, før vi beskriver hver av dem for seg, og til slutt tar med litt om Metode 3. Vi har gitt modellen som i (2.1) på den generelle formen:

$$(7.1) \quad \mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$$

der som før \mathbf{y} er $N \times 1$, \mathbf{X} er $N \times p$ med rang $q < p$, \mathbf{b} er $p \times 1$ og \mathbf{e} er $N \times 1$. Vi har i kapittel 5 (f.eks. (5.4)) definert kvadratsumreduksjonen til parametrene i modellen som:

$$(7.2) \quad R(\mathbf{b}) = \tilde{\mathbf{b}}' \mathbf{X}' \mathbf{y}$$

der $\tilde{\mathbf{b}}$ er en hvilken som helst løsning av $\mathbf{X}' \mathbf{X} \tilde{\mathbf{b}} = \mathbf{X}' \mathbf{y}$. En liten utvidelse av notasjonen er at nå (og i resten av notatet) er argumentet til R en vektor med alle aktuelle parametre, i motsetning til i kapittel 5 der vi brukte α som symbol for alle α_i osv. Notasjonen $\tilde{\mathbf{b}}$ er brukt her for å innbefatte både en

hvilken som helst løsning \mathbf{b}^0 når ligningene ikke har full rang, og løsningen $\hat{\mathbf{b}}$ når rangen er full (som f.eks. i vanlig regresjon), for å understreke at formen i (7.2) er definert slik enten rangen er full eller ikke. Hvis man så deler opp \mathbf{X} og \mathbf{b} slik at $\mathbf{X} = [\mathbf{X}_1 : \mathbf{X}_2]$ og $\mathbf{b}' = [\mathbf{b}_1' : \mathbf{b}_2']$ vil den reduserte modellen $\mathbf{y} = \mathbf{X}_2 \mathbf{b}_2 + \mathbf{e}$ ha en tilhørende kvadratsumreduksjon som er

$$(7.3) \quad R(\mathbf{b}_2) = \tilde{\mathbf{b}}_2' \mathbf{X}_2' \mathbf{y}$$

der $\tilde{\mathbf{b}}_2$ er en vilkårlig løsning til $\mathbf{X}_2' \mathbf{X}_2 \tilde{\mathbf{b}}_2 = \mathbf{X}_2' \mathbf{y}$. Da har vi også at

$$(7.4) \quad R(\mathbf{b}_1 | \mathbf{b}_2) = R(\mathbf{b}_1, \mathbf{b}_2) - R(\mathbf{b}_2)$$

Metode 1

Dette er metoden som benytter den generaliserte inversen, og der det er lettest å se nytten av de lineær-algebraiske resultatene vi har sett på. Den er gitt ved flg. oppskrift:

- (a) Man tar utgangspunkt i modellen fra (7.1) som ikke har full rang: $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$
- (b) Kvadratsumreduksjonen for den fulle modellen er: $R(\mathbf{b}) = \mathbf{b}^{0'} \mathbf{X}' \mathbf{y}$ der $\mathbf{X}' \mathbf{X} \mathbf{b}^0 = \mathbf{X}' \mathbf{y}$. Her kan restriksjoner benyttes for å finne en \mathbf{b}^0 , men resultatet er uavhengig av hvilke restriksjoner som velges.
- (c) En redusert modell (i forhold til den fulle modellen i (a)) er gitt ved $\mathbf{y} = \mathbf{X}_2 \mathbf{b}_2 + \mathbf{e}$
- (d) Kvadratsumreduksjonen til den reduserte modellen er da gitt ved $R(\mathbf{b}_2) = \mathbf{b}_2^{0'} \mathbf{X}_2' \mathbf{y}$, der \mathbf{b}_2^0 er vilkårlig løsning av $\mathbf{X}_2' \mathbf{X}_2 \mathbf{b}_2^0 = \mathbf{X}_2' \mathbf{y}$. Her kan man benytte restriksjoner for å finne \mathbf{b}_2^0 som i (b), og igjen er resultatet uavhengig av hvilke restriksjoner som er benyttet.

I punkt (b) er restriksjonene slike vi som nevnt i innledningen brukte i avsnitt 5.4 for å finne en \mathbf{b}^0 . Der satte vi alle parametrene bortsett fra interaksjonsleddene = 0. Det viktige her er at hadde vi brukt andre restriksjoner hadde vi fått samme resultat på kvadratsumreduksjonen $R(\mathbf{b})$. Grunnen til dette er at $R(\mathbf{b})$ kan skrives på formen $R(\mathbf{b}) = \mathbf{b}^{0'} \mathbf{X}' \mathbf{y} = (\mathbf{G} \mathbf{X}' \mathbf{y})' \mathbf{X}' \mathbf{y} = \mathbf{y} \mathbf{X} \mathbf{G} \mathbf{X}' \mathbf{y}$ og fra Teorem 4.2 husker vi at $\mathbf{X} \mathbf{G} \mathbf{X}'$ er invariant til \mathbf{G} (og dermed \mathbf{b}^0). Altså uansett hvilken \mathbf{b}^0 vi bruker får vi samme $R(\mathbf{b})$.

Metode 2

Nå skal vi se hva som skjer når vi bruker restriksjoner på den overparametriserte modellen direkte, slik at den får full rang.

- (a) Man tar utgangspunkt i modellen fra (7.1), velger restriksjoner slik at man får en ny modell, gitt ved: $\mathbf{y} = \mathbf{X}^* \mathbf{b}^* + \mathbf{e}$. Restriksjonene som her er valgt brukes så videre, i motsetning til metode 1, der forskjellige restriksjoner kan brukes i (a) og i (d).

- (b) Kvadratsumreduksjonen for den fulle modellen med restriksjoner beregnes som:
 $R(\mathbf{b}^*) = \hat{\mathbf{b}}^{*\prime} \mathbf{X}^{*\prime} \mathbf{y}$, der $\hat{\mathbf{b}}^* = (\mathbf{X}^{*\prime} \mathbf{X}^*)^{-1} \mathbf{X}^{*\prime} \mathbf{y}$ siden den nye modellen har full rang.
- (c) Den nye parametervektoren \mathbf{b}^* deles i to slik at $\mathbf{b}^{*\prime} = [\mathbf{b}_1^{*\prime}; \mathbf{b}_2^{*\prime}]$ og en redusert modell spesifiseres som: $\mathbf{y} = \mathbf{X}_2^* \mathbf{b}_2^* + \mathbf{e}$.
- (d) Kvadratsumreduksjonen til den reduserte modellen er da gitt ved $R(\mathbf{b}_2^*) = \hat{\mathbf{b}}_2^{*\prime} \mathbf{X}_2^{*\prime} \mathbf{y}$, der
 $\hat{\mathbf{b}}_2^* = (\mathbf{X}_2^{*\prime} \mathbf{X}_2^*)^{-1} \mathbf{X}_2^{*\prime} \mathbf{y}$

Hovedforskjellen mellom denne metoden og forrige metode er at selv om $R(\mathbf{b}) \equiv R(\mathbf{b}^*)$ er dette ikke nødvendigvis tilfellet for den reduserte modellen, slik at $R(\mathbf{b}_2)$ ikke nødvendigvis er lik $R(\mathbf{b}_2^*)$.

Metode 3

Modeller med full rang er grundig beskrevet i [8]. Som nevnt i innledningen vil vi her nøye oss med å beskrive forbindelsen med de modellene vi har sett på, og så presentere vanlige hypoteser med parametere fra full rang modeller, for å gjøre tolkning av hypotesene lettere. Modellene vi nå omtaler er gitt ved ligningen:

$$(7.5) \quad y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$$

der μ_{ij} er populasjonsgjennomsnitt, som kan tenkes organisert i rader og kolonner som i en 2-veis tabell. Ved å innføre restriksjoner, som kan betraktes som lineære sammenhenger mellom μ_{ij} -ene får man de overparametriserte modellene vi har betraktet før. Uten noen restriksjoner tilsvarer (7.5) vår gjennomgangsmodell, en 2-veis kryssklassifisering med interaksjon som i (4.3).

Modellen gitt ved ligningene:

$$(7.6) \quad \begin{aligned} y_{ijk} &= \mu_{ij} + \varepsilon_{ijk} \\ \mu_{ij} - \mu_{i'j} - \mu_{ij'} + \mu_{i'j'} &= 0, \quad i \neq i', j \neq j' \end{aligned}$$

som er lik den i (7.5), men som har en lineær sammenheng (restriksjon) mellom μ_{ij} -ene tilsvarer en 2-veis kryssklassifisering uten interaksjon. At denne restriksjonen betyr at det ikke er interaksjon kan vi se hvis vi sammenligner med avsnitt 5.4 der vi drøftet hvordan man testet interaksjonsleddet. Man kan også se det ved et enkelt eksempel, f.eks. den 2×2 tabellen i avsnitt 6.2 med $\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$, $n_{11} = n_{22} = 2$ og $n_{12} = n_{21} = 1$. Da blir restriksjonen ekvivalent med $\gamma_{11} - \gamma_{21} = \gamma_{12} - \gamma_{22}$, som igjen er ekvivalent med $\gamma_{12} - \gamma_{11} = \gamma_{22} - \gamma_{21}$, men når dette er oppfylt kan man lage nye variable for radeffekt og kolonneeffekt lik $\tilde{\alpha}_1 = \alpha_1 + \gamma_{11}$, $\tilde{\alpha}_2 = \alpha_2 + \gamma_{12}$, $\tilde{\beta}_1 = \beta_1$ og $\tilde{\beta}_2 = \beta_2 + (\gamma_{12} - \gamma_{11})$. Da får vi den nye modellen gitt ved $\mu_{ij} = \mu + \tilde{\alpha}_i + \tilde{\beta}_j$, $i, j = 1, 2$ som er uten interaksjon og gir samme verdier som opprinnelige modellen med restriksjonen $\gamma_{11} - \gamma_{21} = \gamma_{12} - \gamma_{22}$.

Hvis vi lager en modell med flg. restriksjoner:

$$(7.7) \quad \begin{aligned} y_{ijk} &= \mu_{ij} + \varepsilon_{ijk} \\ \mu_{ij} - \mu_{i'j} - \mu_{ij'} + \mu_{i'j'} &= 0, \quad i \neq i', j \neq j' \\ \mu_{.j} - \mu_{.j'} &= 0, \quad j \neq j' \end{aligned}$$

vil dette tilsvare en 1-veis klassifisering. Den ene restriksjonen er som i (7.6) som betyr ingen interaksjon og den andre restriksjonen er at det ikke er kolonne-effekt (siden summen over radene av populasjonsgjennomsnittene er lik for forskjellige kolonner). Med samme eksempel som ovenfor får restriksjonen formen $\mu_{11} + \mu_{21} = \mu_{12} + \mu_{22}$. Setter vi inn for modellen uten interaksjon (den andre restriksjonen) $\mu_{ij} = \mu + \alpha_i + \beta_j, i, j = 1, 2$, får vi at denne restriksjonen reduseres til $\beta_1 = \beta_2$ som nettopp betyr at det ikke er forskjell på kolonnene og vi har dermed en 1-veis klassifisering. Denne modellen kan selvfølgelig skrives lettere som $y_{ij} = \mu_i + \varepsilon_{ij}$, men (7.7) er med for å illustrere sammenhengen mellom modellen (7.5) og modeller vi tidligere har omtalt. Tilsvarende 1-veis klassifisering for kolonner fås ved å bytte ut j med i i (7.7).

Eksempel

Et eksempel som illustrerer forskjellene mellom Metode 1 og Metode 2 tar utgangspunkt i gjennomgangsmodellen vår fra (4.3), 2-veis kryssklassifisering med interaksjon.

Metode 1: Restriksjonene som gjør løsningen enklest, nemlig å sette alle parametrene bortsett fra interaksjonene lik 0, gir oss fra (5.16) at

$$(7.8) \quad R(\mu, \alpha, \beta, \gamma) = \sum_i \sum_j n_{ij} \bar{y}_{ij}^2$$

Som nevnt ovenfor ville alle mulige restriksjoner gi oss samme resultat. Hvis vi så tenker oss den reduserte modellen $y_{ijk} = \mu + \alpha_i + \varepsilon_{ijk}$ har vi fra (5.3) og (5.4) at vi velger $\mu^0 = 0$ og får

$$(7.9) \quad R(\mu, \alpha) = \sum_i n_i \bar{y}_{i..}^2$$

(vi betrakter nå radene i 2-veis modellen, derfor summeres det over j slik at n_i erstattes av n_i og $\bar{y}_{i..}$ erstattes av $\bar{y}_{i..}$). På tilsvarende måte kan vi finne $R(\mu, \beta)$, $R(\mu, \alpha, \beta)$. Med denne metoden er, som vi husker fra (5.20) $R(\mu, \gamma) = R(\mu, \alpha, \gamma) = R(\mu, \beta, \gamma) = R(\mu, \alpha, \beta, \gamma)$ og grunnen til dette var at rommet utspent av kolonnene til hovedeffektene hver for seg er inneholdt i rommet utspent av kolonnene til interaksjonsleddet.

Metode 2:

For å demonstrere metode 2 velger vi $a = 2$ (2 rader), $b = 3$ (3 kolonner), $n_{11} = 1$ og alle andre $n_{ij} = 2$ (ubalanserte data). Så velger vi f.eks. flg. restriksjoner for å oppnå full rang: $\alpha_1 = \beta_1 = \gamma_{11} = \gamma_{1j} = 0$, for alle i, j . Da blir normallingningene gitt ved:

$$(7.10) \quad \begin{pmatrix} 11 & 6 & 4 & 4 & 2 & 2 \\ 6 & 6 & 2 & 2 & 2 & 2 \\ 4 & 2 & 4 & 0 & 2 & 0 \\ 4 & 2 & 0 & 4 & 0 & 2 \\ 2 & 2 & 2 & 0 & 2 & 0 \\ 2 & 2 & 0 & 2 & 0 & 2 \end{pmatrix} \begin{pmatrix} \hat{\mu}^* \\ \hat{\alpha}_2^* \\ \hat{\beta}_2^* \\ \hat{\beta}_3^* \\ \hat{\gamma}_{22}^* \\ \hat{\gamma}_{23}^* \end{pmatrix} = \begin{pmatrix} y_{\dots} \\ y_{2..} \\ y_{.2} \\ y_{.3} \\ y_{22} \\ y_{23} \end{pmatrix}$$

som har full rang. Her legger vi merke til at alle parametre som er valgt lik 0 er fjernet fra ligningene, og også de tilhørende rader og kolonner (til sammen 6 stykker). Nå kan $\hat{\mu}^*, \dots, \hat{\gamma}_{23}^*$ finnes ved å inverttere matrisen på venstre side, og deretter kan $R(\mu^*, \alpha^*, \beta^*, \gamma^*)$ finnes ved punkt (b) som:

$$(7.11) \quad R(\mu^*, \alpha^*, \beta^*, \gamma^*) = \hat{\mu}^* y_{\dots} + \hat{\alpha}_2^* y_{2..} + \hat{\beta}_2^* y_{.2} + \hat{\beta}_3^* y_{.3} + \hat{\gamma}_{22}^* y_{22} + \hat{\gamma}_{23}^* y_{23}.$$

Da kan det vises at $R(\mu, \alpha, \beta, \gamma) = R(\mu^*, \alpha^*, \beta^*, \gamma^*)$. For å sammenligne hva vi får for den reduserte modellen i (7.9), beregnes $R(\mu^*, \alpha^*)$ ved å sette $\beta_2^* = \beta_3^* = \gamma_{22}^* = \gamma_{23}^* = 0$ i normalligningene i (7.10). Da blir de nye normalligningene:

$$(7.12) \quad \begin{pmatrix} 11 & 6 \\ 6 & 6 \end{pmatrix} \begin{pmatrix} \hat{\mu}^* \\ \hat{\alpha}_2^* \end{pmatrix} = \begin{pmatrix} y_{\dots} \\ y_{2..} \end{pmatrix}$$

som også har full rang. Disse kan nå løses for $\hat{\mu}^*$ og $\hat{\alpha}_2^*$ (mrk. at disse ikke er de samme estimatene som i (7.11)) og dette gir flg. $R(\mu^*, \alpha^*)$:

$$(7.13) \quad R(\mu^*, \alpha^*) = \hat{\mu}^* y_{\dots} + \hat{\alpha}_2^* y_{2..}$$

Igjen kan det vises at fra (7.9) er $R(\mu, \alpha) = R(\mu^*, \alpha^*)$ (mrk. at $\alpha' = (\alpha_1, \alpha_2)$, mens $\alpha^{*'} = \alpha_2^*$). På tilsvarende måte kan vi finne $R(\mu^*, \beta^*)$, $R(\mu^*, \alpha^*, \beta^*)$, $R(\mu^*, \alpha^*, \gamma^*)$, $R(\mu^*, \beta^*, \gamma^*)$. Den viktige forskjellen fra Metode 1 er at nå er $R(\mu^*, \alpha^*, \gamma^*) \neq R(\mu^*, \alpha^*, \beta^*, \gamma^*)$ og tilsvarende

$R(\mu^*, \beta^*, \gamma^*) \neq R(\mu^*, \alpha^*, \beta^*, \gamma^*)$ og grunnen til dette er at nå er ikke kolonnerommet til hver enkelt ovedeffekt inneholdt i kolonnerommet til interaksjonen.

Diskusjon

Noen punkter om sammenligning mellom Metode 1 og Metode 2, vanlige hypoteser under modell (4.3) og (7.5) (Metode 3) for enklere tolkning, er tatt med nedenfor (fra [7]).

Punkt 1

I Metode 2 er ikke $R(\beta_2^*)$ alltid definert, og hvis den er definert er den ikke nødvendigvis unik, den kan variere med restriksjonene. Her har Metode 1 en fordel, for der er alltid $R(\beta_2)$ definert og uavhengig av restriksjonene. Hvis man i Metode 2 f.eks. lager restriksjonene $\mu = \alpha_i = \beta_j = 0$ for alle i

og j er verken $R(\mu^*, \alpha^*)$, $R(\mu^*, \beta^*)$, eller $R(\mu^*, \alpha^*, \beta^*)$ definert. Som et eksempel på $R(\beta_2^*)$ som varierer med restriksjonene kan vi se på $R(\mu^*, \alpha^*, \gamma^*)$. Denne vil variere med restriksjonene gitt nedenfor, som er vanlige restriksjoner under denne modellen:

$$(7.14) \quad \sum_i \alpha_i = \alpha = \beta = \gamma_i = \gamma_j = 0$$

eller

$$(7.15) \quad \alpha_i = \beta_j = \gamma_{ij} = \gamma_{il} = 0 \text{ for alle } i, j$$

eller

$$(7.16) \quad \sum_i n_i \alpha_i = \sum_j n_j \beta_j = \sum_i n_{ij} (\alpha_i + \gamma_{ij}) = \sum_j n_{ij} (\beta_j + \gamma_{ij}) = 0$$

Punkt 2

Som beskrevet på slutten av Metode 2 er $R(\beta_2^*)$ ikke alltid lik $R(\beta_2)$ og Metode 2 er ikke spesialtilfelle av Metode 1. Hvis modellen fra (4.3) er utgangspunkt er $R(\mu, \alpha) = R(\mu^*, \alpha^*)$, $R(\mu, \beta) = R(\mu^*, \beta^*)$ og $R(\mu, \alpha, \beta) = R(\mu^*, \alpha^*, \beta^*)$, forutsatt at alle er definert. Derimot er $R(\mu, \alpha, \gamma) \neq R(\mu^*, \alpha^*, \gamma^*)$ og $R(\mu, \beta, \gamma) \neq R(\mu^*, \beta^*, \gamma^*)$, og dette skyldes som nevnt ovenfor at i Metode 2 spenner ikke kolonnerommet til interaksjonen ut kolonnerommet til hver av hovedeffektene. Metodene er ikke spesialtilfeller av hverandre fordi vi kan i begge tilfeller velge samme restriksjoner, og likevel vil de samme forskjellene som nevnt ovenfor gjelde.

Punkt 3

I verken Metode 1, eller Metode 2 er det klart hvilke hypoteser som testes. Både $R(\beta_2 | \beta_1)$ og $R(\beta_2^* | \beta_1^*)$ er brukt i ANOVA tabeller som utgangspunkt for testing av hypoteser om parametre, men det er ikke alltid intuitivt hvilke hypoteser som testes av hvilke kvadratsummer, og rikelig anledning til å misforstå hva som testes (tolkning av hypotesen). De vanlige hypotesene som er relevante for modellene i (4.3) og (7.5) er vist i Tabell 9 nedenfor. Vi har antatt for enkelhets skyld at alle $n_{ij} > 0$ (selv om ikke alle hypotesene i Tabellen forutsetter dette). Generelt er $i' \neq i$ og $j' \neq j$ (bortsett fra der de er brukt som indeks i summetegn). Tabell 10 viser sammenhengen mellom hypotesene og de forskjellige $R(\)$ leddene. De fleste hypotesene i Tabell 9 med modell (4.3) som utgangspunkt kjenner vi igjen fra avsnitt 5.4, der vi drøftet denne modellen detaljert. Hypotese H2 i Tabell 9, ser vi fra Tabell 10 testes ved $R(\alpha | \mu)$ for modell (4.3) og er samme hypotese som i (5.27). H3 i Tabell 9 testes ved $R(\alpha | \mu, \beta)$ og er den samme som i (5.29). H2 og H3 er begge hypoteser for radeffekten, og helt tilsvarende er H6 og H7 for kolonneeffekten. H9 er testen for interaksjon og er den samme som (5.31) (litt mer presist beskrevet der, og med den viktig bemerkningen angående tolkningen som står drøftet under). Å finne tilhørende hypoteser for reduserte modeller av (4.3) er rett frem. Hvis man f.eks. dropper γ_{ij} -leddene fra H2 i Tabell 9 får vi hypotesen testet ved $R(\alpha | \mu)$ for en 2-veis kryssklassifisering uten interaksjon, slik som fremstilt helt på slutten av avsnitt 5.3.

	Hypotese	Modell (7.5)	Modell (4.3)
Rad-effekt	H1	$\bar{\mu}_i = \bar{\mu}_{i'}$	$\alpha_i + \frac{1}{b} \sum_j \gamma_{ij} = \alpha_{i'} + \frac{1}{b} \sum_j \gamma_{i'j}$
	H2	$\sum_j \frac{n_{ij}}{n_i} \mu_{ij} = \sum_j \frac{n_{i'j}}{n_i} \mu_{i'j}$	$\alpha_i + \sum_j \frac{n_{ij}}{n_i} \beta_j + \sum_j \frac{n_{ij}}{n_i} \gamma_{ij} = \alpha_{i'} + \sum_j \frac{n_{i'j}}{n_i} \beta_j + \sum_j \frac{n_{i'j}}{n_i} \gamma_{i'j}$
	H3	$\sum_j n_{ij} \mu_{ij} = \sum_{i'} \sum_j \frac{n_{ij} n_{i'j}}{n_j} \mu_{i'j}$	$(n_i - \sum_j \frac{n_{ij}^2}{n_j}) \alpha_i + \sum_j (n_{ij} - \frac{n_{ij}^2}{n_j}) \gamma_{ij}$ $= \sum_{i' \neq i} (\sum_j \frac{n_{ij} n_{i'j}}{n_j}) \alpha_{i'} + \sum_{i' \neq i} \sum_j (\frac{n_{ij} n_{i'j}}{n_j}) \gamma_{i'j}$
	H4	$\mu_{i1} = \mu_{i'1}$	$\alpha_i - \alpha_{i'} + \gamma_{i1} - \gamma_{i'1} = 0$
Kolonne-effekt	H5	$\bar{\mu}_j = \bar{\mu}_{j'}$	$\beta_j + \frac{1}{a} \sum_i \gamma_{ij} = \beta_{j'} + \frac{1}{a} \sum_i \gamma_{ij'}$
	H6	$\sum_i \frac{n_{ij}}{n_j} \mu_{ij} = \sum_i \frac{n_{ij'}}{n_j} \mu_{ij'}$	$\beta_j + \sum_i \frac{n_{ij}}{n_j} \alpha_i + \sum_i \frac{n_{ij}}{n_j} \gamma_{ij} = \beta_{j'} + \sum_i \frac{n_{ij'}}{n_j} \alpha_i + \sum_i \frac{n_{ij'}}{n_j} \gamma_{ij'}$
	H7	$\sum_i n_{ij} \mu_{ij} = \sum_{j'} \sum_i \frac{n_{ij} n_{ij'}}{n_i} \mu_{ij'}$	$(n_j - \sum_i \frac{n_{ij}^2}{n_i}) \beta_j + \sum_i (n_{ij} - \frac{n_{ij}^2}{n_i}) \gamma_{ij}$ $= \sum_{j' \neq j} (\sum_i \frac{n_{ij} n_{ij'}}{n_i}) \beta_{j'} + \sum_{j' \neq j} \sum_i (\frac{n_{ij} n_{ij'}}{n_i}) \gamma_{ij'}$
	H8	$\mu_{1j} = \mu_{1j'}$	$\beta_j - \beta_{j'} + \gamma_{1j} - \gamma_{1j'} = 0$
Interaksjon	H9	$\mu_{ij} - \mu_{i'j} - \mu_{ij'} + \mu_{i'j'} = 0$	$\gamma_{ij} - \gamma_{i'j} - \gamma_{ij'} + \gamma_{i'j'} = 0$

Tabell 9: Vanlige ANOVA hypoteser

$R()$	Hypotese
$R(\alpha \mu) = R(\alpha^* \mu^*)$	H2
$R(\beta \mu) = R(\beta^* \mu^*)$	H6
$R(\alpha \mu, \beta) = R(\alpha^* \mu^*, \beta^*)$	H3
$R(\beta \mu, \alpha) = R(\beta^* \mu^*, \alpha^*)$	H7
$R(\gamma \mu, \alpha, \beta) = R(\gamma^* \mu^*, \alpha^*, \beta^*)$	H9
$R(\alpha^* \mu^*, \beta^*, \gamma^*)$ med (7.14)	H1
$R(\alpha^* \mu^*, \beta^*, \gamma^*)$ med (7.15)	H4
$R(\alpha^* \mu^*, \beta^*, \gamma^*)$ med (7.16)	H2
$R(\beta^* \mu^*, \alpha^*, \gamma^*)$ med (7.14)	H5
$R(\beta^* \mu^*, \alpha^*, \gamma^*)$ med (7.15)	H8
$R(\beta^* \mu^*, \alpha^*, \gamma^*)$ med (7.16)	H6

Tabell 10: Sammenheng mellom hypoteser og $R()$ ledd

Hvis man så i tillegg dropper β_j -leddene får man hypotesen testet ved $R(\alpha|\mu)$ fra (5.6) som tester likheten av α_i -ene i en 1-veis klassifisering. På denne måten ser vi at Tabell 9 omfatter alle hypotesene vi har drøftet i kapittel 5.

Fra Tabell 10 ser vi at $R(\)$ notasjonen ikke er presis når det gjelder å beskrive hypotesene som testes. Man kan være fristet til å tolke $R(\alpha|\mu)$ upresist som det å teste α og se bort fra β og γ , $R(\alpha|\mu, \beta)$ som å teste α , se bort fra γ og justere for β , og $R(\alpha|\mu, \beta, \gamma)$ som å teste α og justere for β og γ . Men vi vet at i Metode 1 er $R(\alpha|\mu, \beta, \gamma) = 0$ og i Metode 2 ser vi fra Tabell 10 at denne tester forskjellige hypoteser avhengig av hvilken restriksjon som brukes. F.eks. hvis restriksjon (7.14) brukes testes H1 som inneholder γ_{ij} -ledd så den er avhengig av disse, dermed er "å justere for" ikke i betydningen å fjerne fra hypotesen, og misvisende språkbruk i dette tilfellet. H1 lar seg ikke teste i Metode 1 pga. at den tilhørende kvadratsummen ikke lar seg generere. Når man benytter $R(\)$ notasjonen er det viktig at det fremgår både hvilken metode som brukes og ev. hvilke restriksjoner (Metode 2). Forfatterne av [7] konkluderer med at $R(\)$ leddene er først og fremst en måte (algoritme for) å beregne kvadratsummer på. Det avgjørende er hvilke hypoteser som passer, når man har bestemt seg for det, kan $R(\)$ -leddene brukes for å beregne de korrekte kvadratsummene, med de omtalte begrensninger.

Punkt 4

Noen av hypotesene i Tabell 9 er vanskelig å tolke, men å betrakte dem med to forskjellige modeller som utgangspunkt (4.3) og (7.5), kan være til hjelp. Forbindelsen mellom dem er at ved å sette $\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$ i hypotesene under modell 7.5 (kolonnen til venstre) fås hypotesene under modell (4.3) (kolonnen til høyre).

Hvis vi ser på H1, er denne lett å tolke i μ_{ij} notasjonen fra modell (7.5). Den tester likhet mellom 2 gjennomsnittlige populasjonsgjennomsnitt for rader, når gjennomsnittet beregnes over kolonnene. Fordelen med denne hypotesen er at den er uavhengig av antall observasjoner i hver celle (n_{ij}), så fremt $n_{ij} > 0$. Dette er i tråd med at man skal kunne spesifisere en hypotese før man utfører forsøket, og da vet man ikke nødvendigvis hvor mange observasjoner det er i hver celle (planter som dør i løpet av forsøket o.l.). De andre hypotesene ser vi er avhengig av antallet, og hvis dette antallet ikke representerer antallet i populasjonen for denne kategorien kan det gjøre hypotesen uinteressant. Under modell (4.3) ser vi at den tester likhet mellom to radnivåer med nærvær av gjennomsnittlig interaksjon.

H2 ser vi også sammenligner 2 gjennomsnittlige populasjonsgjennomsnitt for rader, men nå er gjennomsnittet veiet med vektor lik andelen av observasjoner i en celle i forhold til i hele raden. Hvis andelene stemmer med populasjonens fordeling av observasjoner er det en fornuftig hypotese. Ser vi på hvordan hypotesen er formulert under modell (4.5), ser vi at den tester likhet mellom to radnivåer i nærvær av veiede gjennomsnitt av kolonnenivåer og interaksjoner. Hvis vi betrakter eksempelet fra avsnitt 6.2 der vi demonstrerte hva som skjer når dataene er ubalansert med et enkelt talleksempel fra en 2×2 tabell og uten interaksjon med $n_{11} = n_{22} = 2$, $n_{12} = n_{21} = 1$, får vi her at hypotesen tester $H: \alpha_1 - \alpha_2 + \frac{1}{3}(\beta_1 - \beta_2) = 0$ som var nettopp det vi fikk i ligning (6.12) da vi så på differansen mellom de marginale gjennomsnittene. Den assosierte kvadratsummen er $R(\alpha|\mu)$ og her stemmer språkbruken at denne tester α når man ser bort fra β . Som beskrevet i punkt 3 må man være forsiktig

med den muntlige tolkningen av $R(\cdot)$ notasjonen, når man snakker om å justere for en effekt, men dette er kun farlig når betingingen (de effektene bak $|\cdot$) skjer på en effekt som inneholder effekten foran $|\cdot$ (i den forstand at kolonnerommet omfatter den andre effektens kolonnerom).

H3 er vanskelig å tolke. Den er avhengig av antall observasjoner i cellene på en komplisert måte. Den er lettest på formen fra modell (7.5). Ser vi på eksempelet i avsnitt 6.2 igjen (uten interaksjon), får vi ved innsetting at H3 tester hypotesen: $H: \mu_{11} + \mu_{12} = \mu_{21} + \mu_{22}$ som igjen ved innsetting av $\mu_{ij} = \mu + \alpha_i + \beta_j$ gir oss hypotesen $H: 2\alpha_1 + (\beta_1 + \beta_2) = 2\alpha_2 + (\beta_1 + \beta_2)$ som blir $H: \alpha_1 = \alpha_2$ altså teste likhet mellom radnivåene. Den assosierte kvadratsummen er fra Tabell 10 $R(\boldsymbol{\alpha}|\boldsymbol{\mu},\boldsymbol{\beta})$ så her stemmer det også å tolke denne som å teste $\boldsymbol{\alpha}$ og justere for $\boldsymbol{\beta}$. Den har altså justert innblandingen av kolonne-effekten som vi omtalte på slutten av kapittel 6.

H4 er en spesiell hypotese som bare omfatter en kolonne, og som fremkommer kun ved Metode 2 og restriksjonen fra (7.15), og er ikke så anvendelig.

Resten av hypotesene (med unntak av H9 som er omtalt i andre sammenhenger) er tilsvarende disse, men for kolonner i stedet for rader.

8. SAS GLM - kvadratsummer for enhver anledning

I SAS er det mange prosedyrer som kan analysere modellene som vi har beskrevet i dette notatet. Noen er generelle og er ment å dekke mange situasjoner, som f.eks. REG og GLM, mens andre er mer spesialiserte, som f.eks. ANOVA, VARCOMP, NESTED. De spesialiserte prosedyrene er ofte lettere å forstå, de bruker mindre lagringsplass og beregningstid som kan være avgjørende for store datasett, men har et lite anvendelsesområde. I de fleste tilfellene er det overlapp mellom prosedyrene, slik at de kan brukes til mer enn det de først og fremst er tenkt for, men det er viktig å kjenne begrensningene. De generelle prosedyrene, på sin side, er trygge å bruke siden de passer for mange anvendelser, men er vanskeligere å sette seg inn i. Vi skal se nærmere på den generelle prosedyren GLM, fordi dette er den som utnytter mest av teorien vi har sett på, og ved å forstå denne, forstår man også det grunnleggende i de andre prosedyrene.

Samtlige av prosedyrene nevnt ovenfor beregner flere forskjellige typer kvadratsummer, som passer for forskjellige situasjoner, og forskjellige tester. REG og GLM er hovedprosedyrene for regresjon (REG) og forskjellige typer variansanalyse (GLM). De overlapper, men REG har masse diagnose verktøy for regresjon som GLM ikke har. GLM, på sin side kan også brukes til regresjon, men er først og fremst laget for variansanalyse med "fixed" (som vi har sett på) og "random" effekter. Hvis effektene er "random", så er det varianskomponentene som estimeres og dette kan gjøres mer effektivt i VARCOMP, som også har muligheter for "fixed" effekter, men færre muligheter enn GLM. ANOVA kan bare brukes for variansanalyse med "fixed" effekter hvis dataene er balanserte, for da gjelder en del forenklingsformler som gjør beregningene raskere vha matriseregning (se kap. 6), men hvis denne brukes for ubalanserte data blir svarene feil. GLM kan beregne fire forskjellige typer kvadratsummer som er ment å dekke de fleste aktuelle analyser og hypoteser. Om f.eks. dataene er balansert eller ubalansert spiller ingen rolle for GLM (balanserte data er som nevnt før best tenkt på som et spesialtilfelle av ubalanserte data). For å forstå de forskjellige typene kvadratsummer er teorien fra foregående kapitler nødvendig, og vi skal igjen se fordelene ved å beskrive metodene med lineær algebra. Materialet i dette kapittelet er en hentet fra [1],[2] og [8].

8.1 Forskjellige typer kvadratsummer i SAS GLM

GLM (forkortelse for Generelle Lineære Modeller) er som nevnt ovenfor først og fremst beregnet for variansanalyse i de modellene som vi har beskrevet i tidligere kapitler. Den må ikke forveksles med forkortelsen til "Generaliserte Lineære Modeller" som betegner en større klasse av modeller, der modeller for lineær regresjon, logistisk / poisson regresjon, variansanalyse o.a. er spesialtilfeller. SAS har egne prosedyrer for f.eks. logistisk regresjon.

GLM beregner, hvis ønskelig opptil fire forskjellige kvadratsummer, kalt Type I, Type II, Type III, og Type IV. For noen situasjoner er alle disse forskjellige og opphav til forskjellige F -tester så det er viktig å forstå forskjellene på dem. Den enkleste modellen som samtidig er komplisert nok til å få frem de viktige forskjellene er en 2-veis kryssklassifisering med interaksjon som er vår gjennomgangsmodell fra (4.3) og gitt ved: $E(y_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_{ij}$. Tabellen nedenfor viser sammenhengen mellom de fire forskjellige kvadratsumtypene og $R(\cdot)$ notasjonen vi har brukt tidligere. I en 2-veis kryssklassifisering med interaksjon betegnes effektene i SAS som A , B og $A * B$.

I flg. situasjoner er noen av kvadratsumtypene like:

$$(8.1) \quad \begin{array}{ll} I = II = III = IV & \text{Ved balanserte data} \\ II = III = IV & \text{Uten interaksjonsledd} \\ III = IV & \text{Data med ingen tomme celler} \end{array}$$

Som (8.1) viser er det ingen forskjell på de forskjellige kvadratsummene når man har balanserte data, så da kan man like godt bruke prosedyren ANOVA. For modeller uten interaksjonsledd er det bare Type I som skiller seg ut, og for data med observasjoner i alle kombinasjoner av i og j er Type III det samme som Type IV.

FAKTOR	TYPE I (SEKVENSIELL)	TYPE II (JUSTERT)	TYPE III (MED RESTR.)	TYPE IV (TOMME CELLER)
A	$R(\alpha \mu)$	$R(\alpha \mu, \beta)$	$R(\alpha^* \mu^*, \beta^*, \gamma^*)$	data-avh.
B	$R(\beta \mu, \alpha)$	$R(\beta \mu, \alpha)$	$R(\beta^* \mu^*, \alpha^*, \gamma^*)$	data-avh.
$A * B$	$R(\gamma \mu, \alpha, \beta)$	$R(\gamma \mu, \alpha, \beta)$	$R(\gamma \mu, \alpha, \beta)$	$R(\gamma \mu, \alpha, \beta)$

Tabell 11: Sammenheng mellom forskjellige typer kvadratsummer i SAS og $R(\cdot)$ notasjon

Tabell 9 og 10 (i kapittel 7) viser hvilke hypoteser de respektive kvadratsummene er assosiert med.

8.2 Estimerbare funksjoner i SAS GLM

Fra kapittel 4 vet vi at det er en nær sammenheng mellom testbare hypoteser og estimerbare funksjoner. Siden SAS GLM sine forskjellige kvadratsumtyper er assosiert med forskjellige hypoteser, har de også korresponderende estimerbare funksjoner, som SAS også har valgt å dele inn i de samme omtalte fire typene. I utskriften av de forskjellige kvadratsumtypene gis også utskrift av formen på de tilhørende estimerbare funksjonene, som er en hjelp for å se hvordan hypotesen som kvadratsummen tester ser ut. Dette gjelder også Tabell 9 og 10 i kapittel 7. For å illustrere viser vi to eksempler på slike utskrifter, begge på Type I kvadratsummer, hentet fra [8]. Utskriften av den estimerbare funksjonen består av to kolonner, den ene som kalles "Effect" som skal illustrere de forskjellige faktorene med deres nivåer, slik at "Intercept" betegner μ , A_1, A_2, \dots betegner de forskjellige nivåene til faktoren α og tilsvarende for de andre faktorene. Den andre kolonnen kalles "Coefficient" og er for

hver linje angitt med en kombinasjon av ett eller flere av symbolene $L1, L2, L3, \dots$ med fortegn og eventuell koeffisient foran.

To eksempler

Først ser vi på en enkel 1-veis klassifisering med 3 nivåer for en faktor α . Da vil utskriften av en Type I kvadratsum foruten selve verdien av kvadratsummen også ha flg. tabell:

EFFECT	COEFFICIENT
Intercept	
$A1$	$L2$
$A2$	$L3$
$A3$	$-L2 - L3$

Tabell 12: Estimerbar funksjon for Type 1 kvadratsum i 1-veis klassifisering med 3 nivåer

Fra denne utskriften fås den estimerbare funksjonen som en sum av produktet mellom elementene i kolonne 1 og kolonne 2. Hvis vi kaller den estimerbare funksjonen for f , bruker l_1, l_2, l_3, \dots i stedet for $L1, L2, L3, \dots$ og vanlig modell notasjon gir Tabell 12 flg. estimerbare funksjon:

$$(8.2) \quad f = l_2 \alpha_1 + l_3 \alpha_2 - (l_2 + l_3) \alpha_3$$

Så ser vi på en 2-veis kryssklassifisering uten interaksjon med 2 nivåer for en faktor α (rader), og 3 nivåer for β (kolonner) og med flg. n_{ij} verdier og som gir SAS utskriften i tabellen nederst:

	n_{ij}		n_i	
	3	3	1	7
	2	1	.	3
n_j	5	4	1	10

Tabell 13: Antall observasjoner i 2-veis kryssklassifisering uten interaksjon

EFFECT	COEFFICIENT
Intercept	
$A1$	$L2$
$A2$	$-L2$
$B1$	$-0.2381 L2$
$B2$	$0.0952 L2$
$B3$	$0.1429 L2$

Tabell 14: Estimerbar funksjon for Type 1 kvadratsum i 2-veis kryssklassifisering uten interaksjon

På samme måte som i eksempelet ovenfor vil den estimerbare funksjonen i Tabell 14 være gitt ved:

$$(8.3) \quad f = l_2 \alpha_1 - l_2 \alpha_2 - 0.2381 l_2 \beta_1 + 0.0952 l_2 \beta_2 + 0.1429 l_2 \beta_3$$

De to estimerbare funksjonene i (8.2) og (8.3) har to viktige egenskaper:

NB 1:

For alle verdier av l_1, l_2, \dots i f er den en estimerbar funksjon. Hvis vi f.eks. i (8.2) setter inn de to lineært uavhengige settene med $\mathbf{I}'_1 = (l_2 \ l_3) = (1 \ -1)$ og $\mathbf{I}'_2 = (l_2 \ l_3) = (1 \ 0)$ fås de to lineært uavhengige estimerbare funksjonene:

$$(8.4) \quad f_1 = \alpha_1 - \alpha_2 \quad \text{og} \quad f_2 = \alpha_1 - \alpha_3$$

I den estimerbare funksjonen i (8.3) for 2-veis klassifiseringen er det bare en l_2 (l_1 er i begge disse eksemplene assosiert med μ , men μ er justert for i disse kvadratsummene og kommer derfor ikke inn i den estimerbare funksjonen, se nedenfor). Med $l_2 = 1$ blir den estimerbare funksjonen i (8.3) lik:

$$(8.5) \quad f_1 = \alpha_1 - \alpha_2 - 0.2381\beta_1 + 0.0952\beta_2 + 0.1429\beta_3$$

NB 2:

Antall forskjellige symboler l_1, l_2, \dots i en estimerbar funksjon f er lik antall frihetsgrader, f.eks. r_A , i kvadratsummen som kan skrives som en kvadratisk form $\mathbf{y}'\mathbf{A}\mathbf{y}$ (vi husker fra avsnitt 5.3 at alle kvadratsummene vi har sett på kan skrives som en slik kvadratisk form) og som er assosiert med f (r_A vil da være rangen til \mathbf{A}). De to eksemplene ovenfor har begge kvadratsummen (Type I) $R(\boldsymbol{\alpha}|\mu)$ assosiert med de respektive estimerbare funksjonene som er gitt ved:

$$(8.6) \quad \begin{aligned} \text{For 1-veis klassifisering i eks.: } R(\boldsymbol{\alpha}|\mu) &= \sum_{i=1}^3 n_i (\bar{y}_i - \bar{y}_{..})^2 \quad \text{og} \quad r_A = 2 \\ \text{For 2-veis klassifisering i eks.: } R(\boldsymbol{\alpha}|\mu) &= \sum_{i=1}^2 n_i (\bar{y}_{i..} - \bar{y}_{...})^2 \quad \text{og} \quad r_A = 1 \end{aligned}$$

Vi kjenner igjen $R(\boldsymbol{\alpha}|\mu)$ for 1-veis klassifisering fra ligning (5.6) og for 2-veis klassifisering fra ligning (5.17) ved å huske at $R(\boldsymbol{\alpha}|\mu) = R(\mu, \boldsymbol{\alpha}) - R(\mu)$ og samme omforming som i (5.6).

Hypoteser fremstilt ved de estimerbare funksjonene

Etter å ha uttrykt de estimerbare funksjonene er det enkelt å uttrykke de tilhørende hypotesene som kvadratsummene tester. Hvis vi kaller vektoren med $r_A = r$ forskjellige symboler l_1, l_2, \dots, l_r i f for $\mathbf{I}' = (l_1 \quad l_2 \quad \dots \quad l_r)$, kan r_A lineært uavhengige forskjellige \mathbf{I}' brukes til å gi r_A forskjellige (maksimalt antall lineært uavhengige) estimerbare funksjoner f_1, f_2, \dots, f_r . Da er hypotesene assosiert med kvadratsummene og korresponderende f gitt ved:

$$(8.7) \quad H: f_i = 0 \quad \text{for} \quad i = 1, \dots, r_A$$

Anvendt på de to eksemplene gir dette, fra (8.4) med $r_A = 2$, flg. hypotese for 1-veis klassifisering:

$$(8.8) \quad \text{For 1-veis eks.} \quad H: \begin{cases} \alpha_1 - \alpha_2 = 0 \\ \alpha_1 - \alpha_3 = 0 \end{cases}, \quad \text{eller} \quad H: \alpha_1 = \alpha_2 = \alpha_3$$

som vi husker er testen som assosiert med $R(\boldsymbol{\alpha}|\mu)$ (fra slutten av avsnitt 5.2). Her er det viktig å legge merke til at med andre valg av \mathbf{I}'_1 og \mathbf{I}'_2 som er lineært uavhengige, fås den samme hypotesen. For eksempelet med 2-veis kryssklassifisering uten interaksjon er $r_A = 1$. Med $l_2 = 1$ som i (8.5) og ved å sette $f_1 = 0$ fås flg. hypotese for dette tilf.:

$$(8.9) \quad \text{For 2-veis eks.} \quad H: \alpha_1 - \alpha_2 - 0.2381\beta_1 + 0.0952\beta_2 + 0.1429\beta_3 = 0$$

Denne er ikke så lett å kjenne igjen, men er den samme hypotesen som er beskrevet på slutten av avsnitt 5.3 (og som H2 i Tabell 9, når man ser bort fra interaksjonsleddet). Der er hypotesen gitt som $H: \alpha_i + \sum_j \frac{n_{ij}}{n_i} \beta_j$ like $\forall i$. For å se at dette er samme setter vi inn verdier fra Tabell 13 og får hypotesen $H: \alpha_1 + (3\beta_1 + 3\beta_2 + \beta_3)/7 = \alpha_2 + (2\beta_1 + \beta_2)/3$ som ved å trekke sammen gir hypotesen $H: \alpha_1 - \alpha_2 + (\frac{3}{7} - \frac{2}{3})\beta_1 + (\frac{3}{7} - \frac{1}{3})\beta_2 + \frac{1}{7}\beta_3 = 0$ som er den samme som (8.9). Også her er det slik at med andre valg av l_2 fås samme hypotesen. Dette viser at sammenhengen mellom koeffisientene i den estimerbare funksjonen og modellen, eller verdiene av n_{ij} , ikke nødvendigvis er så enkel, selv i dette ganske enkle eksempelet. Nedenfor skal vi se nærmere på en generell sammenheng.

Generelt

Nå skal vi nevne et generelt resultat som er nyttig når det gjelder å uttrykke det maksimale antallet lineært uavhengige hypoteser (som er det samme som det maksimale antallet lineært uavhengige estimerbare funksjoner) ut fra en kvadratisk form med en matrise som ikke nødvendigvis har full rang. Siden dette er et generelt resultat er det nyttig og brukes i SAS (i hvert fall er dette tolkningen av det som gjøres).

Utgangspunktet er som før at vi har $\mathbf{y} \sim N(\mathbf{X}\mathbf{b}, \sigma^2\mathbf{I})$. Da vil en F -observator på formen

$\mathbf{y}'\mathbf{A}\mathbf{y}/r_A\hat{\sigma}^2$ (ikke-sentralt F -fordelt), der \mathbf{A} er symmetrisk (husk for usymmetriske matriser finnes en tilhørende symmetrisk matrise slik at den kvadratiske formen blir den samme), idempotent, har dimensjon $N \times N$ og rang r_A , teste hypotesen:

$$(8.10) \quad H: \mathbf{T}'\mathbf{X}\mathbf{b} = \mathbf{0}$$

der \mathbf{T}' har full rad-rang lik r_A og med rader som er egenvektorene som svarer til enhets-egenverdien til \mathbf{A} . Dette er r_A lineære hypoteser om parametrene i modellen, som er det vi vil ha, selv om dimensjonen på \mathbf{A} er N .

For å begrunne dette ser vi først at når observatoren er F -fordelt som beskrevet er kravene for dette som i Teorem 3.3 - 3.5 og da er blant annet \mathbf{A} idempotent. Da gjelder (generelt for idempotente matriser, se [4] s 302) at \mathbf{A} kan uttrykkes som $\mathbf{A} = \mathbf{T}\mathbf{T}'$ for $\mathbf{T}'\mathbf{T} = \mathbf{I}$ der kolonnene til \mathbf{T} er de r_A ortogonale egenvektorene til \mathbf{A} tilsvarende de r_A egenverdiene som er lik 1 (en idempotent matrise \mathbf{A} av dimensjon N og rang r_A har r_A egenverdier lik 1 og $N - r_A$ egenverdier lik 0, se [4] s 321). Videre er hypotesen som F -observatoren opprinnelig tester og som tidligere omtalt er det uttrykket som gjør at ikke-sentralitetsparameteren blir 0, gitt ved:

$$(8.11) \quad H: \mathbf{b}'\mathbf{X}'\mathbf{A}\mathbf{X}\mathbf{b} = \mathbf{0}$$

Det er flere måter å forenkle denne hypotesen på. Ved å bruke at kvadratiske former som er kvadratsummer, her $\mathbf{y}'\mathbf{A}\mathbf{y}$, kjennetegnes ved at matrisen \mathbf{A} er symmetrisk og ikke-negativt definit (se [9], s 230). Da kan \mathbf{A} skrives som $\mathbf{A} = \mathbf{P}\mathbf{P}'$ for en reell matrise \mathbf{P}' med full rad-rang r_A (ved full-rang faktorisering, se [4] s 206). Dermed kan (8.11) skrives som $\mathbf{b}'\mathbf{X}'\mathbf{A}\mathbf{X}\mathbf{b} = \mathbf{b}'\mathbf{X}'\mathbf{P}\mathbf{P}'\mathbf{X}\mathbf{b} = \mathbf{0}$ som impliserer $\mathbf{P}'\mathbf{X}\mathbf{b} = \mathbf{0}$, se [4] s 62. Da er hypotesen i (8.11) ekvivalent med $H: \mathbf{P}'\mathbf{X}\mathbf{b} = \mathbf{0}$. Dette er pga. full rad-rang til \mathbf{P}' nettopp r_A lineært uavhengige hypoteser om parametrene, slik som vi er ute etter, men problemet er at \mathbf{P}' kan være vanskelig å finne. Et alternativ kan da være å se at $\mathbf{P}'\mathbf{X}\mathbf{b} = \mathbf{0}$ impliserer $\mathbf{P}\mathbf{P}'\mathbf{X}\mathbf{b} = \mathbf{0}$ og siden $\mathbf{A} = \mathbf{P}\mathbf{P}'$ gir dette hypotesen $H: \mathbf{A}\mathbf{X}\mathbf{b} = \mathbf{0}$. Dette er en enklere form enn (8.11), men det er N hypoteser om parametrene hvor bare r_A av disse er lineært uavhengige og det kan være

vanskelig å plukke ut disse. Men ved å bruke egenvektormatrisen \mathbf{T} beskrevet ovenfor har vi at denne siste formen på hypotesen kan skrives $H: \mathbf{T}\mathbf{T}'\mathbf{X}\mathbf{b} = \mathbf{0}$ som ved premultiplisering med \mathbf{T}' og ved at $\mathbf{T}'\mathbf{T} = \mathbf{I}$ gir hypotesen $H: \mathbf{T}'\mathbf{X}\mathbf{b} = \mathbf{0}$ som i (8.10). Dette er også r_A lineært uavhengige hypoteser om parametrene, og en metode som alltid virker (å finne egenverdier og egenvektorer er relativt enkelt).

Hvis vi nå betrakter den generelle testbare hypotesen $H: \mathbf{K}'\mathbf{b} = \mathbf{0}$ fra avsnitt 4.4 (med $\mathbf{m} = \mathbf{0}$, dette er en begrensning i GLM), med de samme kravene som der at \mathbf{K}' har dimensjon $s \times p$ og full rad-rang s kan vi gjøre bruk av resultatet over. For en vilkårlig ikke-singulær matrise \mathbf{L} med dimensjon s , er denne generelle hypotesen ekvivalent med:

$$(8.12) \quad H: \mathbf{L}\mathbf{K}'\mathbf{b} = \mathbf{0}$$

og siden den vilkårlige \mathbf{L} kan betraktes som s lineært uavhengige vektorer \mathbf{I}' med dimensjon $1 \times s$ kan hypotesen i (8.12) også skrives som

$$(8.13) \quad H: \mathbf{I}'\mathbf{K}'\mathbf{b} = \mathbf{0} \quad \text{for vilkårlig } s \text{ lineært uavhengige vektorer } \mathbf{I}' \text{ med lengde } s$$

Uttrykket $\mathbf{I}'\mathbf{K}'\mathbf{b}$ i (8.13) er en estimerbar funksjon (siden $\mathbf{K}'\mathbf{b}$ er estimerbar). Ved å definere denne funksjonen som f

$$(8.14) \quad f = \mathbf{I}'\mathbf{K}'\mathbf{b}$$

og bruke s lineært uavhengige vektorer $\mathbf{I}'_1, \dots, \mathbf{I}'_s$ for å beregne s forskjellige verdier av f lik $f_i = \mathbf{I}'_i\mathbf{K}'\mathbf{b}$ for $i = 1, \dots, s$ får vi at hypotesen i (8.13) kan skrives som

$$(8.15) \quad H: \begin{cases} f_1 = 0 \\ \vdots \\ f_s = 0 \end{cases}$$

som nettopp er på formen fra (8.7). Den estimerbare funksjonen i (8.14) er nettopp den type estimerbar funksjon som utskriften i SAS GLM presenterer, som hver kvadratsum $\mathbf{y}'\mathbf{A}\mathbf{y}$, med tilhørende F -observator, er assosiert med. Men vi har samtidig at en slik F -observator tester hypotesen i (8.10) så da må den estimerbare funksjonen i (8.14) kunne skrives

$$(8.16) \quad f = \mathbf{I}'\mathbf{T}'\mathbf{X}\mathbf{b}$$

med de omtalte egenskapene til \mathbf{T}' (\mathbf{K}' må kunne skrives som $\mathbf{K}' = \mathbf{T}'\mathbf{X}$ siden F -observatoren tester begge formene av hypotesen). Dette er et eksempel på en \mathbf{T}' som vi i avsnitt 4.4 diskuterte ("beviste") eksistensen av. Da var hovedhensikten å utlede F -observatoren for den generelle hypotesen. I (4.33) satte vi opp den passende kvadratiske formen (med $\mathbf{m} = \mathbf{0}$) $Q = \mathbf{b}'\mathbf{K}'\mathbf{K}(\mathbf{K}'\mathbf{G}\mathbf{K})^{-1}\mathbf{K}'\mathbf{b}$, med tilhørende normalfordelt $\mathbf{K}'\mathbf{b}$ og slik at matrisen $(\mathbf{K}'\mathbf{G}\mathbf{K})^{-1}$ er idempotent og symmetrisk (se avsnitt 4.4). Den har dimensjon $s \times s$ og $r_A = s$. Da ser vi at ved resultatet i (8.10) vil de s egenvektorene til $(\mathbf{K}'\mathbf{G}\mathbf{K})^{-1}$ kunne settes sammen til en \mathbf{T}' og teste hypotesen $H: \mathbf{b}'\mathbf{K}'\mathbf{K}(\mathbf{K}'\mathbf{G}\mathbf{K})^{-1}\mathbf{K}'\mathbf{b} = \mathbf{0}$ ved den enklere formen $H: \mathbf{T}'\mathbf{K}'\mathbf{b} = \mathbf{0}$. Her er \mathbf{T}' ikke-singulær slik at denne hypotesen er ekvivalent med $H: \mathbf{K}'\mathbf{b} = \mathbf{0}$ som var utgangspunktet (8.10) bekrefter dermed vår utledning i avsnitt 4.4). Her er det viktig å legge

merke til at hypotesen $H: \mathbf{b}'\mathbf{K}(\mathbf{K}'\mathbf{G}\mathbf{K})^{-1}\mathbf{K}\mathbf{b} = \mathbf{0}$ ikke automatisk medfører $H: \mathbf{K}'\mathbf{b} = \mathbf{0}$, selv om det motsatte er tilfellet.

Ytterligere forenkling er mulig og vanlig å benytte (som i SAS GLM). Vi ser at hypotesen i (8.12) også lar seg formulere som $H: \mathbf{L}\mathbf{S}^{-1}\mathbf{K}'\mathbf{b} = \mathbf{0}$ for vilkårlige ikke-singulære matriser \mathbf{L} og \mathbf{S} . Derfor er den vanligste formen på den estimerbare funksjonen i (8.16) i stedet

$$(8.17) \quad f = \mathbf{l}'\mathbf{S}^{-1}\mathbf{T}'\mathbf{X}\mathbf{b}$$

der \mathbf{S} kun er valgt for å forenkle uttrykket, slik at koeffisienten foran de s første elementene i \mathbf{b} kun er ett enkelt element i $\mathbf{l}' = (l_1 \ \dots \ l_s)$. Utskriften i SAS består av parametrene i modellen (\mathbf{b}) og ved siden av disse de tilhørende koeffisientene i (8.17) der elementene i \mathbf{l}' er skrevet som $L1, L2, L3, \dots$ til sammen s forskjellige elementer.

Eksempelene fortsatt

Bruker vi denne metoden på eksemplene ovenfor, får vi for 1-veis klassifiseringen med $n_1 = 3$, $n_2 = 2$ og $n_3 = 2$:

$$(8.18) \quad R(\boldsymbol{\alpha}|\mu) = \sum_{i=1}^3 n_i (\bar{y}_i - \bar{y}_{..})^2 = \mathbf{y}'\mathbf{A}\mathbf{y} \text{ for } \mathbf{A} = \begin{pmatrix} \bar{\mathbf{J}}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{J}}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \bar{\mathbf{J}}_2 \end{pmatrix} - \bar{\mathbf{J}}_7 \text{ og } r_A = 2$$

der $\bar{\mathbf{J}}_n$ er en kvadratisk matrise av dimensjon n med alle elementene lik $\frac{1}{n}$. Da kan det vises at en passe \mathbf{T}' med $r_A = 2$ rader som hver er egenvektorer til \mathbf{A} som korresponderer med enhetseigenverdien til \mathbf{A} er:

$$\mathbf{T}' = \begin{pmatrix} 2 & 2 & 2 & 0 & 0 & -3 & -3 \\ 0 & 0 & 0 & 2 & 2 & -2 & -2 \end{pmatrix}$$

Dette gir:

$$\mathbf{T}'\mathbf{X}\mathbf{b} = \begin{pmatrix} 0 & 6 & 0 & -6 \\ 0 & 0 & 4 & -4 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}$$

og fra (8.17) får vi følgende estimerbare funksjon:

$$f = (l_1 \ l_2) \begin{pmatrix} 6 & 0 \\ 0 & 4 \end{pmatrix}^{-1} \begin{pmatrix} 0 & 6 & 0 & -6 \\ 0 & 0 & 4 & -4 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} = l_2\alpha_1 + l_3\alpha_2 - (l_2 + l_3)\alpha_3$$

som vi kjenner igjen fra (8.2) med \mathbf{S} valgt som beskrevet. Vi så i (8.8) at denne estimerbare funksjonen er egnet for å uttrykke en hypotese på den generelle formen i (8.12).

Helt tilsvarende blir det i vår 2-veis kryssklassifisering med antall observasjoner som i Tabell 13:

$$R(\boldsymbol{\alpha}|\mu) = \sum_{i=1}^2 n_i (\bar{y}_{i..} - \bar{y}_{...})^2 = \mathbf{y}'\mathbf{A}\mathbf{y} \text{ for } \mathbf{A} = \begin{pmatrix} \bar{\mathbf{J}}_7 & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{J}}_3 \end{pmatrix} - \bar{\mathbf{J}}_{10} \text{ og } r_{\mathbf{A}} = 1$$

Her kan det vises at flg. egenvektor til \mathbf{A} korresponderer med den ene egenverdien som er lik 1:

$$\mathbf{t}' = (3 \ 3 \ 3 \ 3 \ 3 \ 3 \ 3 \ -7 \ -7 \ -7)$$

som gir

$$\mathbf{t}'\mathbf{X}\mathbf{b} = (0 \ 21 \ -21 \ -5 \ 2 \ 3)(\mu \ \alpha_1 \ \alpha_2 \ \beta_1 \ \beta_2 \ \beta_3)'$$

og dermed fra (8.17)

$$\begin{aligned} f &= l_2(21)^{-1} (0 \ 21 \ -21 \ -5 \ 2 \ 3)(\mu \ \alpha_1 \ \alpha_2 \ \beta_1 \ \beta_2 \ \beta_3)' \\ &= l_2 \left[(\alpha_1 - \alpha_2) - \frac{5}{21}\beta_1 + \frac{2}{21}\beta_2 + \frac{3}{21}\beta_3 \right] \end{aligned}$$

som stemmer med (8.9).

Vi husker fra avsnitt 5.3 at vi der begynte med å spesifisere en ønsket hypotese, f.eks. likhet blant α -ene i en slik 2-veis modell som her i eksempelet, og ved å sette opp en \mathbf{K}' matrise for dette utledet at den kvadratiske formen Q i den generelle F -observatoren i (4.35) var lik kvadratsummen $R(\boldsymbol{\alpha}|\mu, \boldsymbol{\beta})$ (tilsvarende er mulig for $R(\boldsymbol{\alpha}|\mu)$ som her). Som vi husker betød dette at $R(\boldsymbol{\alpha}|\mu, \boldsymbol{\beta})$ (som ikke er en teoretisk størrelse, men bestemt av dataene) tester, som teller i F -observatoren, hvorvidt den tilhørende teoretiske størrelsen $\mathbf{b}'\mathbf{K}(\mathbf{K}'\mathbf{G}\mathbf{K})^{-1}\mathbf{K}'\mathbf{b}$, som er en ikke-sentralitets parameter, er forskjellig fra 0. Hvis denne parameteren er positiv er fordelingen forskjøvet til høyre fra en sentral F -fordeling. Så hvis test-observatoren, sammenlignet med en sentral F -fordeling gir forkastning, betyr det at dataene tyder på at et slikt (eller mer ekstremt) resultat er svært usannsynlig hvis fordelingen er sentral, altså at fordelingen antakelig er ikke-sentral og $\mathbf{b}'\mathbf{K}(\mathbf{K}'\mathbf{G}\mathbf{K})^{-1}\mathbf{K}'\mathbf{b} \neq 0$, som igjen medfører at $\mathbf{K}'\mathbf{b} \neq \mathbf{0}$ og dermed forskjell på nivåene i α .

I en programutskrift som i SAS er kvadratsummene gitt, og hypotesene de tester fremkommer ved de estimerbare funksjonene.

8.3 I beste mening

Hvis vi tenker oss hypotesen H : *Alle manualer er gode og fullstendige beskrivelser skrevet i beste mening*, kan bakgrunnen for dette notatet betraktes som en test av denne. Slik all signifikanstesting er basert på tenkte gjentatte forsøk og manualene på trimrommet krever repetisjoner for å gi fremgang, gjelder dette også for tekniske manualer. Ved å lese det samme flere ganger, kanskje lære litt ny teori i mellomtiden, øker forståelsen. Vi skrev i innledningen at det kunne være svært vanskelig å skjønne hva SAS GLM gjør og hvorfor (utfra manualbeskrivelsen). Gjennomgang av stoffet hittil burde gjøre det lettere å forstå hvorfor fremstillingen i manualen er slik den er og vi finner ikke grunn til å forkaste hypotesen ovenfor. I avsnitt 8.2 og 8.3 prøvde vi å forklare hva SAS GLM gjør fra en mer teoretisk synsvinkel, mens i dette avsnittet skal vi se litt på manualens fremstilling og oppsummerende om de forskjellige typene kvadratsummer med tilhørende estimerbare funksjoner.

Siden SAS innskrenker hypotesene til $H: \mathbf{K}'\mathbf{b} = \mathbf{0}$ er denne, som vist i (8.12), ekvivalent med $H: \mathbf{L}\mathbf{K}'\mathbf{b} = \mathbf{0}$ for en ikke-singulær \mathbf{L} med rang (dimensjon) s . Denne siste formen representerer som nevnt s lineært uavhengige estimerbare funksjoner satt lik 0, ved å fylle \mathbf{L} med symboler i stedet for tall. Hvis man så lar programmet finne enkleste mulige form på koeffisientene fås formen i (8.17) som er den formen som SAS GLM presenterer. SAS GLM gjør ikke forskjell på hypotesen og de tilhørende estimerbare funksjonene, pga. nevnte sammenheng. Manualen beskriver at kvadratsummer for alle testbare hypoteser $\mathbf{L}\mathbf{b} = \mathbf{0}$ kan beregnes ved uttrykket:

$$(8.19) \quad SS(H_0: \mathbf{L}\mathbf{b} = \mathbf{0}) = (\mathbf{L}\mathbf{b}^0)' (\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}')^{-1} (\mathbf{L}\mathbf{b}^0)$$

Her kjenner vi igjen \mathbf{L} som den vi har kalt \mathbf{K}' , men som fremdeles har betydningen av å inneholde de estimerbare funksjonene (som rader), så notasjonen er tilsvarende (med 1 matrise i stedet for 2). Setter man inn verdier for elementene i \mathbf{L} , slik at det blir s lineært uavhengige rader fås en verdi på høyresiden som er kvadratsummen Q fra (4.35), eller altså et $R(\)$ ledd som i SAS GLM betegnes med $SS(\)$.

Den første utskriften i SAS GLM er den såkalte generelle estimerbare funksjonen og det er nettopp den estimerbare funksjonen som korresponderer med kvadratsummen $R(\)$ med alle parametrene, altså i vår 2-veis kryssklassifisering med interaksjon er det $R(\mu, \alpha, \beta, \gamma)$. Den er gitt på samme form som eksemplene på $R(\alpha|\mu)$ i avsnitt 8.2, med parametrene langs en kolonne, og de respektive koeffisientene og uttrykkene i $L1, L2, \dots$ i kolonnen ved siden av. Vi husker fra avsnitt 5.3 at alle $R(\)$ leddene er en F -observator med tilhørende hypotese som kan testes, som vi nå vet fremkommer ved å sette den estimerbare funksjonen lik 0. Den generelle estimerbare funksjonen i SAS GLM utskriften gir en oversikt over alle funksjoner som er estimerbare ved å sette inn verdier for koeffisientene. Hvis man f.eks. er ute etter en estimerbar funksjon (hypotese) kun for en av faktorene må man velge verdier på de andre koeffisientene slik at de andre faktorene elimineres. Der det ikke lar seg gjøre betyr det at den funksjonen som involverer kun den ene faktoren ikke er estimerbar (den tilhørende hypotesen ikke er testbar). Med symbol-koeffisient notasjonen er det på denne måten lett å se om en hvilken som helst funksjon som inneholder ønskede parametre er estimerbar (om hypotesen er testbar).

Etter den generelle estimerbare funksjonen kommer utskrift av samme form for de forskjellige faktorene (A , B og $A*B$ i vår 2-veis klassifisering) med de assosierte F -testene nedenfor. Denne utskriften er avhengig av hvilken type (Type I - IV) kvadratsum / estimerbar funksjon man har bedt om å få. En viktig ting å huske angående disse testene er at uansett hvilke verdier man velger for $L1, L2, \dots$, så lenge de forskjellige valgene gir lineært uavhengige estimerbare funksjoner (rader) i \mathbf{L} , vil man få samme testen (dette er pga. (8.12)). Nedenfor følger litt oppsummering om de forskjellige typene kvadratsummer / estimerbare funksjoner.

Type I

Kvadratsummer av Type I er betegnet med "sekvensiell" i Tabell 11 fordi den er bestemt av rekkefølgen på de forskjellige faktorene når de spesifiseres i prosedyren. Hoved-effekter må spesifiseres før interaksjonsledd, og en faktor som inneholder en nestet faktor må komme før den nastede faktoren (vi har ikke sett på nastede faktorer). Tabell 11 viser at rekkefølgen på de spesifiserte faktorene har vært A , B og $A*B$ og de respektive kvadratsummene som beregnes er da $R(\alpha|\mu)$, $R(\beta|\mu, \alpha)$ og $R(\gamma|\mu, \alpha, \beta)$. Hvis rekkefølgen hadde vært B , A og $A*B$ hadde de beregnede kvadratsummene vært $R(\beta|\mu)$, $R(\alpha|\mu, \beta)$ og $R(\gamma|\mu, \alpha, \beta)$. Med den upresise språkbruken fra

kapittel 7 kan vi si at kvadratsummen justerer for de faktorene som kommer før i spesifiseringen, i den forstand at den faktoren som justeres for ikke er med i uttrykket for den assosierte hypotesen. En hypotese for en faktor vil altså generelt inneholde parametre for seg selv og alle faktorer som kommer etter i spesifiseringen. Denne måten å inkludere flere og flere faktorer i modellen ligner litt på variabelseleksjon i vanlig regresjon, men ikke så mye brukt innen variansanalyse. Type I er kanskje best egnet for sammenligning med de andre typene noe som kan gi et slags mål på hvor stor effekten av ubalanse i dataene er, men den er også egnet for noen anvendelser som vi ikke har sett på, som f.eks. noen tester for heterogenitet blant regresjonskoeffisienter (se [1]). Fra Tabell 9 ser vi at de assosierte hypotesene til Type I er avhengig av antall observasjoner i cellene, n_{ij} , og kan derfor være uinteressant.

En fin egenskap ved Type I kvadratsummer er at de er additive, dvs. at de summeres til den totale kvadratsummen. For vår 2-veis klassifisering får vi at

$$R(\alpha|\mu) + R(\beta|\mu, \alpha) + R(\gamma|\mu, \alpha, \beta) = R(\mu, \alpha, \beta, \gamma) - R(\mu) = R(\alpha, \beta, \gamma|\mu),$$

dvs. summen av de tre respektive kvadratsummene gir totalkvadratsummen for modellen, korrigert for μ . Dette er siden

$$R(\alpha|\mu) = R(\mu, \alpha) - R(\mu), \quad R(\beta|\mu, \alpha) = R(\mu, \alpha, \beta) - R(\mu, \alpha),$$

og

$$R(\gamma|\mu, \alpha, \beta) = R(\mu, \alpha, \beta, \gamma) - R(\mu, \alpha, \beta).$$

Den tilhørende estimerbare funksjonen av Type I er beregningsmessig enkel og fremkommer ved radene som ikke er 0-rader i en såkalt "Forward Doolittle" matrise. Denne matrisen er rad-operasjoner på $\mathbf{X}'\mathbf{X}$ (som bevarer rad-rangen), vi husker fra avsnitt 4.3 at dette er en estimerbar funksjon, og er slik at for en 2-veis kryssklassifisering med interaksjon består den av lineært uavhengige rader, en for hver faktor (resten av radene er 0-rader) og slik at med raden for faktor α betegnet med \mathbf{L}_α , blir kvadratsummen assosiert med hypotesen $H: \mathbf{L}_\alpha \mathbf{b} = 0$ (fra 8.19) lik $R(\alpha|\mu)$ som i Tabell 11.

Tilsvarende, hvis raden for faktor β kalles \mathbf{L}_β blir kvadratsummen assosiert med hypotesen

$H: \mathbf{L}_\beta \mathbf{b} = 0$ lik $R(\beta|\mu, \alpha)$ (også som i Tabell 11), og med \mathbf{L}_γ som raden for interaksjonen γ blir kvadratsummen assosiert med hypotesen $H: \mathbf{L}_\gamma \mathbf{b} = 0$ lik $R(\gamma|\mu, \alpha, \beta)$ i Tabell 11. I denne modellen har alle kvadratsummene 1 frihetsgrad og derfor fremkommer alle mulige hypoteser (estimerbare funksjoner) f.eks. for α ved at \mathbf{L} i (8.19) blir radvektoren \mathbf{L}_α (som for β og γ). Det ene eksempelet i avsnitt 8.2 var en slik 1 frihetsgrads hypotese. Det var en 2-veis klassifisering (riktignok uten interaksjon, men det er uten betydning her) hvor vi så i den estimerbare funksjonen i Tabell 14 at det kun inngikk en symbolkoeffisient (L_2) som enklest kan settes =1 og gi hypotesen i (8.9) (andre verdier gir samme hypotese).

For balanserte data i vår 2-veis klassifisering ville Type I estimerbare funksjon for α , bare inneholde parametrene for α og γ , i stedet for sånn som i det ubalanserte tilfellet der β også er med. Som vi husker fra kapittel 6 er i det balanserte tilfellet $R(\alpha|\mu) = R(\alpha|\mu, \beta)$ som vi tolket som at faktor α blir justert for β (i tillegg til μ). I det balanserte tilfellet er alle typene like (ligning 8.1).

Type II

Som Tabell 11 viser er en Type II kvadratsum den samme som Type I i noen tilfeller, men ikke i alle. I vår 2-veis klassifisering er den lik Type I for faktor β , men ikke for α . Den kan kalles

$R(\text{en faktor}|\text{alle andre passende faktorer})$. De respektive kvadratsummene er $R(\alpha|\mu, \beta)$, $R(\beta|\mu, \alpha)$ og $R(\gamma|\mu, \alpha, \beta)$. De justerer (i samme betydning som i Type I) en faktor for andre faktorer, bortsett fra interaksjonsledd der faktoren inngår, eller nestede faktorer som faktoren inneholder. At den ikke kan justere for et interaksjonsledd der den inngår husker vi fra diskusjonen i kapittel 7 (punkt 3, 4) er fordi

at kolonnerommet assosiert med en faktor er et underrom av kolonnerommet assosiert med interaksjonsleddet, så hvis interaksjonen "fjernes" forsvinner også faktoren selv. Så kvadratsummen lar seg ikke generere i den overparametriserte modellen (Metode 1 fra kapittel 7). Den assosierte hypotesen som testes ved Type II er fra Tabell 9 den som kalles H3 (og H7). Vi drøftet under punkt 4 hvordan denne justerer for den andre hoved-effekten sin innblanding som oppstår pga. ubalansen i dataene (se også slutten av kapittel 6). Det er også verdt å merke seg at den justerer for interaksjonsledd der den ikke inngår (når det er flere enn to hoved-effekter), men altså ikke hvis den inngår, og disse presiseringene er det som kan oppsummeres som å justere for "alle andre passende faktorer", eller som i Tabell 11 bare kalles "justert". Der justerer den faktor A for B , og faktor B for A (lik Type I når man spesifiserer faktorene i rekkefølgen A , B og $A * B$). Kvadratsummen for interaksjonsleddet er justert for begge hoved-effektene. I likhet med Type I, er Type II også avhengig av n_{ij} , noe som kan gjøre den uinteressant hvis populasjonsandelene ikke er like celle-andelene (som nevnt i kapittel 7).

Type II kvadratsummer er generelt ikke lengre additive slik Type I er.

Siden Type II ikke kan justere for interaksjon, er denne kvadratsummen best egnet for situasjoner uten interaksjon (hvis man vet at det ikke er interaksjon, eller man finner at interaksjonsleddet ikke er signifikant). Hvis interaksjonsleddet er signifikant kan man bruke Type III i stedet, eller man kan splitte opp analysen, og undersøke ev. forskjeller i den ene faktoren for hvert nivå av den andre.

At en estimerbar funksjon av Type II ikke inneholder andre parametre for en faktor enn for faktoren selv og ev. faktorer den er inneholdt i, gjør at den ligner mer på det balanserte tilfellet enn Type I. Dette er også noe av motivasjonen for å konstruere Type II. Beregningsmessig kan man oppnå Type II ved å gjenta Type I beregningene med varierende rekkefølge på faktorene og samle på den som ikke kommer først (og derfor er justert for den første). I praksis løser algoritmen i SAS GLM dette på en tilsvarende måte, men uten å måtte forandre rekkefølgen.

I en Type II hypotese for en faktor som er inneholdt i en annen faktor (f.eks. hypotese om α i vår 2-veis kryssklassifisering med interaksjon) vil koeffisientene i \mathbf{L} til interaksjonen, som i Type I, også være avhengig av antall observasjoner i cellene, noe som kan gjøre den uinteressant.

Type III

En kvadratsum av Type III tilsvarer Metode 2 i kapittel 7, som går ut på å lage restriksjoner i den overparametriserte modellen slik at den får full rang. Dette er eneste måten å generere kvadratsummer som kan justere en hoved-effekt for interaksjonen den inngår i. Selv om (som nevnt i kapittel 7) det å justere nå ikke betyr å fjerne fra tilsvarende hypotese, så fjernes "deler" av interaksjonen (den delen som ikke er estimerbar). På samme måte som Type II justerer den for andre hoved-effekter, og fra diskusjonen i kapittel 7 (punkt 2) har vi at uten interaksjonsledd blir kvadratsummen lik Type II. Det er fordi $R(\mu, \alpha, \beta) = R(\mu^*, \alpha^*, \beta^*)$, $R(\mu, \alpha) = R(\mu^*, \alpha^*)$ og $R(\mu, \beta) = R(\mu^*, \beta^*)$ slik at

$R(\alpha^* | \mu^*, \beta^*) = R(\alpha | \mu, \beta)$ og $R(\beta^* | \mu^*, \alpha^*) = R(\beta | \mu, \alpha)$ såfremt alle leddene er definert. Dette betyr at det kun er når det er interaksjon med i modellen forskjellen mellom Type II og Type III er av betydning og Type III har altså en slags justering også for interaksjonen. Derfor kan Type III kalles $R(\text{en faktor} | \text{alle andre faktorer})$ og er beregnet å kunne brukes når man f.eks. skal sammenligne nivåene i en faktor, justert for den andre faktoren, men i nærvær av (estimerbar) interaksjon. Kvadratsummen for interaksjonsleddet er justert for begge hoved-effektene som i Type I og Type II.

Type III kvadratsummer er som Type II generelt heller ikke summerbare.

Type III kvadratsummen har den fordel at de assosierte hypotesene ikke er avhengig av n_{ij} og er dermed mer i tråd med det å kunne sette opp en hypotese før man ev. vet hvor mange observasjoner

det er i hver celle, eller hvis man starter med balanserte data og frafall inntreffer underveis er det mest naturlig at ikke hypotesene forandres.

I de fleste tilfellene med ubalanserte data, er formen på den generelle estimerbare funksjonen den samme som hvis dataene hadde vært balansert, som nevnt i eksempel 2.2. Siden Type II estimerbare funksjoner er avhengig av antall observasjoner vil den for en faktor i vår 2-veis klassifisering variere, avhengig av om dataene er balanserte, eller ikke. Type III estimerbare funksjoner varierer altså ikke med antall observasjoner i cellene, slik at når den generelle estimerbare funksjonen er uforandret så skal også de estimerbare funksjonene (testbare hypotesene) være uforandret. Den skal også, som i Type II, ikke inneholde parametre for andre faktorer enn seg selv og faktorer som den inngår i. For å gjøre Type III uavhengig av antall observasjoner gjøres hypotesen assosiert med en faktor ortogonal med hypotesen for faktoren den inngår i. Dette er nemlig noe som kjennetegner hypoteser i balanserte data. Type I og Type II har ikke denne egenskapen, der er ortogonaliteten til stede i balansert tilfelle, men ikke i ubalansert. I vår 2-veis klassifisering skal altså hypotesene assosiert med hver av faktorene være ortogonale på hypotesen assosiert med interaksjonen. Denne egenskapen gjør at Type III er den som ligner mest på det balanserte tilfellet og er noe av tanken bak den.

En estimerbar funksjon av Type III er generert direkte fra den generelle estimerbare funksjonen. Et eksempel på en Type III estimerbar funksjon i vår 2-veis kryssklassifisering med interaksjon følger nedenfor. Antall nivåer i begge faktorene er 2. Den generelle estimerbare funksjonen er

EFFECT	COEFFICIENT
Intercept	$L1$
$A1$	$L2$
$A2$	$L1 - L2$
$B1$	$L4$
$B2$	$L1 - L4$
$AB11$	$L6$
$AB12$	$L2 - L6$
$AB21$	$L4 - L6$
$AB22$	$L1 - L2 - L4 + L6$

Tabell 15: Den generelle estimerbare funksjon i 2-veis kryssklassifisering med interaksjon

Fra denne konstrueres en Type III hypotese for interaksjonen γ ved først å sette $L1 = L2 = L4 = 0$ i Tabell 15. Da får vi:

EFFECT	COEFFICIENT
Intercept	0
$A1$	0
$A2$	0
$B1$	0
$B2$	0
$AB11$	$L6$
$AB12$	$-L6$
$AB21$	$-L6$
$AB22$	$L6$

Tabell 16: Type III estimerbar funksjon for interaksjonen

For å oppnå en Type III estimerbar funksjon for α tas også utgangspunkt i den generelle estimerbare funksjonen. Ved å sette $L1 = L4 = 0$ oppnås å fjerne faktorene μ og β fra den estimerbare funksjonen. Interaksjonen lar seg ikke fjerne uten å fjerne α også. Nå er den estimerbare funksjonen avhengig av $L2$ og $L6$. For å forenkle ytterligere settes $L6 = K \times L2$. Vi ser at samme vektorer kan uttrykkes med denne formen som med $L2$ og $L6$ ved passende valg av K . Dette gjør at den funksjonen bare er avhengig av $L2$ av de opprinnelige symbolkoeffisientene og en konstant. Denne estimerbare funksjonen representerer alle de mulige estimerbare funksjonene uten μ og β . Ved så å bestemme denne konstanten slik at den estimerbare funksjonen er ortogonal til den estimerbare funksjonen for interaksjonen (Tabell 15) fås den endelige formen. Her er dette ved $K = 0.5$. Disse operasjonene er vist nedenfor:

FAKTOR	GENERELL FUNKSJON	$L1 = L4 = 0$	$L6 = K L2$	$K = 0.5$
Intercept	$L1$	0	0	0
$A1$	$L2$	$L2$	$L2$	$L2$
$A2$	$L1 - L2$	$-L2$	$-L2$	$-L2$
$B1$	$L4$	0	0	0
$B2$	$L1 - L4$	0	0	0
$AB11$	$L6$	$L6$	$K L2$	$0.5 L2$
$AB12$	$L2 - L6$	$L2 - L6$	$(1 - K)L2$	$0.5 L2$
$AB21$	$L4 - L6$	$-L6$	$-K L2$	$-0.5 L2$
$AB22$	$L1 - L2 - L4 + L6$	$-L2 + L6$	$(K - 1)L2$	$-0.5 L2$

Tabell 17: Type III estimerbar funksjon for α i 2-veis kryssklassifisering med interaksjon

Det er måten K bestemmes på som avgjør om den estimerbare funksjonen er av Type II eller Type III. Type II har K avhengig av antall observasjoner i cellene, mens Type III bestemmer K slik som ovenfor.

Type IV

Denne typen kvadratsum er konstruert først og fremst for å kunne takle tomme celler, altså at man ikke har observasjoner for noen kombinasjoner av i og j (en eller flere $n_{ij} = 0$). Den er basert på at ut fra mønsteret av tomme celler bestemmer GLM hvilke hypoteser som passer. Det er til dels sammensatte kriterier som ligger bak hvilke hypoteser som velges. Type IV kvadratsummer benytter seg ikke nødvendigvis av alle dataene og de er ikke nødvendigvis unike for et datasett. Vi har valgt å ikke beskrive de forskjellige tilfellene av Type IV, ettersom vi ikke har funnet at de inneholder så mye nytt om modellene vi har sett på. Vi nøyer oss med å henvise til [1] og synes det er tilstrekkelig å huske at dette er en type kvadratsum som takler det å ha tomme celler (som nevnt i kapittel 7 forutsetter noen hypoteser at alle $n_{ij} > 0$). Hvis ingen celler er tomme er denne kvadratsummen lik Type III.

6. Referanser

- [1]: "SAS system for linear models", Littell, R. C., Freund, R. J., Spector, Ph. C.
SAS Series in Statistical Applications, Third Edition, Cary, NC: SAS Institute Inc., 1991
- [2]: "SAS User's Guide: Statistics", Version 5 Edition
Cary, NC: SAS Institute Inc., 1985
- [3]: "Linear Models", Searle, S., R.,
John Wiley & Sons, Inc., 1971, Wiley Classics Library Edition 1997
- [4]: "Matrix algebra useful for statistics", Searle, S., R.,
John Wiley & Sons, Inc., 1982
- [5]: "Test of Hypotheses in Fixed Effects Linear Models", Goodnight, J., H.
SAS Technical Report R-101, Cary, N.C.: SAS Institute Inc., 1978
- [6]: "Methods of Analysis of Linear Models with Unbalanced Data", Speed, F., M., Hocking, R.,
R., and Hackney, O., P.,
Journal of the American Statistical Association, Mach 1978, Volume 73, Number 361
- [7]: "The Use of the R() - Notation with Unbalanced Data", Speed, F., M., and Hocking, R., R.
The American Statistician, February 1976, Vol. 30, No. 1
- [8]: "Linear Models for Unbalanced Data", Searle, Shayle R.,
John Wiley & Sons, Inc., 1987

Apendiks A - Derivasjon m.h.p. vektorer

Den deriverte til en skalar funksjon $f(\mathbf{x})$ av kolonnevektoren \mathbf{x} , mhp $\mathbf{x} = (x_1, \dots, x_n)'$ defineres som kolonnevektoren:

$$(A.1) \quad \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_n} \end{bmatrix}$$

slik at f.eks. den deriverte av skalarproduktet $\mathbf{c}'\mathbf{x}$ for en vilkårlig vektor av konstanter \mathbf{c} mhp vektoren \mathbf{x} blir det samme som den deriverte av skalarproduktet $\mathbf{x}'\mathbf{c}$ (siden $\mathbf{c}'\mathbf{x} = \mathbf{x}'\mathbf{c}$) og lik:

$$(A.2) \quad \frac{\partial}{\partial \mathbf{x}}(\mathbf{c}'\mathbf{x}) = \frac{\partial}{\partial \mathbf{x}}(\mathbf{x}'\mathbf{c}) = \mathbf{c}$$

Tilsvarende blir det å derivere mhp en radvektor:

$$(A.3) \quad \frac{\partial}{\partial \mathbf{x}'}(\mathbf{c}'\mathbf{x}) = \frac{\partial}{\partial \mathbf{x}'}(\mathbf{x}'\mathbf{c}) = \left[\frac{\partial}{\partial \mathbf{x}}(\mathbf{c}'\mathbf{x}) \right]' = \mathbf{c}'$$

For vektorfunksjoner blir utvidelsen slik at når $\mathbf{y} = (y_1, \dots, y_m)'$ er en vektorfunksjon der elementene er funksjoner av \mathbf{x} , defineres det to mulige deriverte, nemlig:

$$(A.4) \quad \frac{\partial \mathbf{y}'}{\partial \mathbf{x}} = \left[\frac{\partial}{\partial \mathbf{x}} y_1, \dots, \frac{\partial}{\partial \mathbf{x}} y_m \right] \text{ og } \frac{\partial \mathbf{y}}{\partial \mathbf{x}'} = \begin{bmatrix} \frac{\partial}{\partial \mathbf{x}'} y_1 \\ \vdots \\ \frac{\partial}{\partial \mathbf{x}'} y_m \end{bmatrix} = \left[\frac{\partial \mathbf{y}'}{\partial \mathbf{x}} \right]'$$

som er matriser av dimensjon $n \times m$ og $m \times n$ henholdsvis. Eksempler på (A.4) er

$$(A.5) \quad \frac{\partial \mathbf{x}}{\partial \mathbf{x}'} = \frac{\partial \mathbf{x}'}{\partial \mathbf{x}} = \mathbf{I}, \quad \frac{\partial}{\partial \mathbf{x}'}(\mathbf{A}\mathbf{x}) = \mathbf{A} \frac{\partial \mathbf{x}}{\partial \mathbf{x}'} = \mathbf{A}, \quad \frac{\partial}{\partial \mathbf{x}}(\mathbf{x}'\mathbf{B}) = \frac{\partial \mathbf{x}'}{\partial \mathbf{x}} \mathbf{B} = \mathbf{B}$$

der \mathbf{A} og \mathbf{B} er matriser som ikke er funksjoner av \mathbf{x} .

Den deriverte av indreproduktet $\mathbf{u}'\mathbf{v}$ med \mathbf{u} og \mathbf{v} som vektorer der hvert element er funksjon av \mathbf{x} , er vha produktregelen definert som:

$$(A.6) \quad \frac{\partial(\mathbf{u}'\mathbf{v})}{\partial \mathbf{x}} = \frac{\partial}{\partial \mathbf{x}} \sum_i u_i v_i = \sum_i \left[\frac{\partial u_i}{\partial \mathbf{x}} v_i + u_i \frac{\partial v_i}{\partial \mathbf{x}} \right] = \frac{\partial \mathbf{u}'}{\partial \mathbf{x}} \mathbf{v} + \frac{\partial \mathbf{v}'}{\partial \mathbf{x}} \mathbf{u}$$

ved formen i (A.4). Dermed kan den deriverte av en kvadratisk form utledes fra (A.6) som:

$$(A.7) \quad \frac{\partial}{\partial \mathbf{x}}(\mathbf{x}'\mathbf{A}\mathbf{x}) = \frac{\partial \mathbf{x}'}{\partial \mathbf{x}} \mathbf{A}\mathbf{x} + \frac{\partial \mathbf{x}'\mathbf{A}'}{\partial \mathbf{x}} \mathbf{x} = \mathbf{A}\mathbf{x} + \mathbf{A}'\mathbf{x}$$

Apendiks B - Positivt definite matriser

At en symmetrisk matrise \mathbf{A} er positivt definit er en egenskap knyttet til den korresponderende kvadratiske formen $\mathbf{x}'\mathbf{A}\mathbf{x}$. Den er positivt definit hvis:

$$(B.1) \quad \mathbf{x}'\mathbf{A}\mathbf{x} > 0 \text{ for alle } \mathbf{x} \neq \mathbf{0}$$

Dersom $\mathbf{x}'\mathbf{A}\mathbf{x} = 0$ for $\mathbf{x} = \mathbf{0}$ og for en annen $\mathbf{x} \neq \mathbf{0}$ er \mathbf{A} positivt semi-definit. Dersom \mathbf{A} er enten positivt definit, eller positivt semi-definit sies den å være ikke-negativt definit. At det bare omtales som en egenskap til symmetriske matriser er pga. at det egentlig er en egenskap til den kvadratiske formen $\mathbf{x}'\mathbf{A}\mathbf{x}$ og dersom \mathbf{A} ikke er symmetrisk finnes det likevel en entydig symmetrisk matrise \mathbf{B} gitt ved $\mathbf{B} = \frac{1}{2}(\mathbf{A} + \mathbf{A}')$ som har den samme kvadratiske formen, altså for \mathbf{A} ikke symmetrisk finnes alltid en symmetrisk \mathbf{B} slik at $\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{x}'\mathbf{B}\mathbf{x}$ for alle \mathbf{x} (se [3], s 34).

Apendiks C - Matrisemultiplikasjon

Forskjellige måter å betrakte produktet av matrisene $\mathbf{A}_{p \times q}$ (dimensjon $p \times q$) og $\mathbf{B}_{q \times r}$ er verdifullt for å fremheve forskjellige egenskaper til resultatet. Hvis vi tenker oss $\mathbf{A}_{p \times q}$ spesifisert ved sine rader og $\mathbf{B}_{q \times r}$ ved sine kolonner er produktet gitt ved:

$$(C.1) \quad \mathbf{AB} = \begin{pmatrix} \text{rad}(\mathbf{A})_1 \\ \vdots \\ \text{rad}(\mathbf{A})_p \end{pmatrix} \mathbf{B} = \begin{pmatrix} \text{rad}(\mathbf{A})_1 \mathbf{B} \\ \vdots \\ \text{rad}(\mathbf{A})_p \mathbf{B} \end{pmatrix}$$

eller tilsvarende:

$$(C.2) \quad \mathbf{AB} = \mathbf{A}(\text{kol}(\mathbf{B})_1 \quad \cdots \quad \text{kol}(\mathbf{B})_r) = (\mathbf{A} \text{kol}(\mathbf{B})_1 \quad \cdots \quad \mathbf{A} \text{kol}(\mathbf{B})_r)$$

Videre hvis elementene i radene til \mathbf{A} er gitt ved $\text{rad}(\mathbf{A})_1 = (a_{11} \quad \cdots \quad a_{1q})$ osv. til $\text{rad}(\mathbf{A})_p = (a_{p1} \quad \cdots \quad a_{pq})$ og kolonne-elementene til \mathbf{B} er

$$\text{kol}(\mathbf{B})_1 = \begin{pmatrix} b_{11} \\ \vdots \\ b_{q1} \end{pmatrix} \quad \cdots \quad \text{kol}(\mathbf{B})_r = \begin{pmatrix} b_{1r} \\ \vdots \\ b_{qr} \end{pmatrix}$$

vil produktet \mathbf{AB} også kunne ses på som:

$$(C.3) \quad \mathbf{AB} = \begin{pmatrix} a_{11} \text{rad}(\mathbf{B})_1 + \cdots + a_{1q} \text{rad}(\mathbf{B})_q \\ \vdots \\ a_{p1} \text{rad}(\mathbf{B})_1 + \cdots + a_{pq} \text{rad}(\mathbf{B})_q \end{pmatrix}$$

eller som

$$(C.4) \quad \mathbf{AB} = (\text{kol}(\mathbf{A})_1 b_{11} + \cdots + \text{kol}(\mathbf{A})_q b_{q1} \quad \cdots \quad \text{kol}(\mathbf{A})_1 b_{1r} + \cdots + \text{kol}(\mathbf{A})_q b_{qr})$$

Dvs. at matriseproduktet kan ses på som en lineærkombinasjon av radene til \mathbf{B} med vektorer fra radene til \mathbf{A} , eller som en lineærkombinasjon av kolonnene til \mathbf{A} med vektorer fra kolonnene til \mathbf{B} . Bevisene for disse resultatene er rett frem.

De sist utgitte publikasjonene i serien Notater

- 2002/28 C. Nordseth og T. Sandnes: FD - Trygd: Dokumentasjonsrapport. Foreløpig uførestønad. 1992-2000. 37s.
- 2002/29 S. Derakhshanfar og T. Sandnes: FD - Trygd: Dokumentasjonsrapport. Økonomisk sosialhjelp. 1992-2000. 36s.
- 2002/30 I. Johansen: Undersøking om foreldrebetaling i barnehagar, januar 2002. 42s.
- 2002/31 T.M. Køber, H. Moafi, E. Rønning og Ø. Sivertstøl: Bruk av forløpsdatabaser i Statistisk sentralbyrå. 60s.
- 2002/32 T.M. Normann: Omnibusundersøkelsen februar/mars 2002. Dokumentasjonsrapport. 37s.
- 2002/33 S. Reid: Bosettingskriteriene i inntektssystemet til kommunene. Erfaringer med overgang til ny beregningsmåte og nye bosettingskriterier, 2002. 43s.
- 2002/34 K.E. Engebretsen, P.E. Gjerdtnet, S. Kristoffersen, P.G. Larssen og J.H. Wang: Mottak og tilrettelegging av SLN-data. 49s.
- 2002/35 D. Rafat: Analyse av sammenheng mellom ektefellers sysselsetting i en familie. 27s.
- 2002/36 A. Bruvoll og T. Bye: En vurdering av avfallspolitikkens bidrag til løsning av miljø- og ressursproblemer. 31s.
- 2002/37 K.I. Bøe: B.R. Joneid: KOSTRA revisjonssystem. Malverk for generelt revisjonssystem - KOSTRA-data. Revidert utgave. 66s.
- 2002/38 N. Arnesen, G. Daugstad, O.E. Hallingstad, E. Skretting Lunde og B.Vold: Kvalitetssikring i KOSTRA. Forslag til dokumentasjonsrutiner med erfaring fra FylkesKOSTRA-helsetjenester, somatikk. 54s.
- 2002/39 H. Moafi: Omlegging av folkehøgskolestatistikk. Overgang til elektronisk rapportering. 31s.
- 2002/40 Ø. Kleven: Mediebrukundersøkelsen 2001. Dokumentasjonsrapport. 43s.
- 2002/41 Ø. Kleven: Samordnet levekårsundersøkelse 2000 - panelundersøkelsen. Dokumentasjonsrapport. 129s.
- 2002/42 L. Solheim: Foreløpige tall i FoB2001 Utvalg, vekter, estimering og usikkerhet. 64s.
- 2002/43 A. Andersen, E. Birkeland, J. Epland og M. I. Kirkeberg: Økonomi og levekår for ulike grupper trygdemottakere 2001. Foreløpig rapport. 214s.
- 2002/44 E.E. Eibak og R. Johannessen: Forventningsindikator - konsumprisene. Mai-november 2002. 16s.
- 2002/45 D. Quang Pham: Konkrete problemer med sesongjustering i SSB. 189s.
- 2002/46 A. Akselsen, G. Dahl og B.R. Joneid: FD - Trygd. Dokumentasjonsrapport. 1992-1997. 48s.
- 2002/47 J. Kristiansen: Visualisering av statistikk. Fra tabell til diagram. 40s.
- 2002/48 A. Finstad: Utslippsfaktorer for benzen. 18s.
- 2002/49 T. Bye, K.M. Heide og E. Holmøy: Transportutvikling i langsiktige fremskrivninger for norsk økonomi. Prosjektnotat til ECON. 30s.
- 2002/50 L. Solheim, M.I. Faldmo og D. Sve: Foreløpige tall i Jordbrukstelling 1999. Dokumentasjon av metoder og produksjon. 68s.
- 2002/51 KOSTRA: Arbeidsgrupperapporter 2002. 219s.
- 2002/52 Årsrapport 2001 Kontaktutvalget for helse- og sosialstatistikk. 35s.
- 2002/53 E. Dalheim: En skjemabasert komplettering av registeret over befolkningens høyeste utdanning - Opplysninger om opplæring, skolegang og utdannig 1999. 59s.