Prof. H. Skjervold.

Med vennlig hilsen fra forfatteren.

ANVENDT "LEAST-SQUARES"-TEORI MED HOVEDVEKTEN LAGT PÅ RANGERING AV POTENSIELLE AVLSDYR I DET PRAKTISKE AVLSARBEIDET

Forelesninger for dr. studenter ved
Institutionen för husdjursförädling
Lantbrukshögskolan

av

Knut Rönningen

[redacted]

Uppsala 1976

ANVENDT "LEAST-SQUARES"-TEORI MED HOVEDVEKTEN LAGT PÅ RANGERING AV POTENSIELLE AVLSDYR I DET PRAKTISKE AVLSARBEIDET

FORELESNINGER FOR DR. STUDENTER VED INSTITUTIONEN FÖR HUSDJURSFÖRÄDLING, LANTBRUKSHÖGSKOLAN

AV

KNUT RÖNNINGEN

UPPSALA 1976

INNHOLDSFORTEGNELSE

I. INNLEDNING

I de siste 10 årene har det skjedd en rask utvikling når det gjelder metoder for rangering av avlsdyr. Spesielt har Cornell-skolen i USA under ledelse av prof. dr. Henderson värt den ledende på dette området. Mye av teorien er nokså komplisert, og i mange arbeider er teorien beskrevet på et usammenhengende og heller vanskelig sett.

Hensikten med dette kompendiet er derfor å summere opp teorien uten å legge hovedvekten på teoretiske bevis. Derimot vil hovedvekten bli lagt på å illustrere ulike metoder ved bruk av nummeriske eksempler. Etter som en har kommet lengst i anvendelse av metodene på storfe (nötkreatur), vil eksemplene selvsagt også få en viss "slagside" i den retningen. Det bör imidlertid understrekes at metodene er generelle, og de kommer sannsynligvis til å bli brukt også på andre dyreslag i fremtiden.

Det er önskelig at studentene har innhentet forkunnskaper motsvarende Searle (1967), Fimland (1972) og deler av Searle (1971) för dette kompendiet leses.

II. KORT REPITISJON OMKRING GENERELLE LINEÄRE MODELLER

A. Definisjon

La den generelle lineäre modellen väre:

$$\underline{y} = X\underline{b} + \underline{e}$$

hvor $\underline{y}$ er en $N \times 1$ vektor av observasjoner

X er en $N \times k$ matrise av konstanter

$\underline{b}$ er en $k \times 1$ vektor av parametrer

$\underline{e}$ er en $N \times 1$ vektor av ukjente elementer

Vi skal så se på noen spesialtilfeller.

B. Regresjonsfunksjon

Ved analyse av en simpel (enkel) regresjonsfunksjon har en fölgende modell:

$$y_i = \alpha + b_1 X_i + e_i \qquad i = 1, 2, \ldots, N$$

Hvis alle observasjonene skrives får en:

$$\begin{aligned} y_1 &= \alpha + b_1 X_1 + e_1 \\ y_2 &= \alpha + b_1 X_2 + e_2 \\ y_3 &= \alpha + b_1 X_3 + e_3 \\ &\vdots \\ y_N &= \alpha + b_1 X_N + e_N \end{aligned}$$

som gir:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & X_3 \\ \vdots & \vdots \\ 1 & X_N \end{bmatrix} \begin{bmatrix} \alpha \\ b_1 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ \vdots \\ e_N \end{bmatrix}$$

Som reduserer til den generelle modell:

$$\underline{y} = X\underline{b} + \underline{e}$$

For å få trening i å benytte matriser er det viktig å ha en viss kjennskap til hvordan ulike matriser ser ut. En vil derfor gå gjennom ulike matrise-operasjoner og vise hvordan ulike matrise-uttrykk ser ut.

$$X'X = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ X_1 & X_2 & X_3 & \cdots & X \end{bmatrix} \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & X_3 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & X_N \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{N} 1^2 & \sum_{i=1}^{N} 1 \cdot X_i \\ \sum_{i=1}^{N} 1 \cdot X_i & \sum_{i=1}^{N} X_i^2 \end{bmatrix}$$

$$= \begin{bmatrix} N & \sum_{i=1}^{N} X_i \\ \sum_{i=1}^{N} X_i & \sum_{i=1}^{N} X_i^2 \end{bmatrix} = \begin{bmatrix} N & X. \\ X. & X.^2 \end{bmatrix}$$

hvor X' er den transponerte ("transpose") X.

Vi har videre at:

$$X'Y = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ X_1 & X_2 & X_3 & \dots & X_N \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N} 1 \cdot y_i \\ \sum_{i=1}^{N} X_i y_i \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N} y_i \\ \sum_{i=1}^{N} X_i y_i \end{bmatrix}$$

Hvert enkelt element av disse matriser skulle väre kjent fra för.

La oss utvide modellen til en <u>multippel</u> regresjon med 3 uavhengige variable:

$$Y_i = \alpha + b_1 X_{i_1} + b_2 X_{i_2} + b_3 X_{i_3} + e_i \quad , \quad i = 1, 2, 3, \dots , N$$

Hvis alle observasjonene skrives, får en:

$$\begin{aligned} y_1 &= \alpha + b_1 X_{11} + b_2 X_{12} + b_3 X_{13} + e_1 \\ y_2 &= \alpha + b_1 X_{21} \quad b_2 X_{22} + b_3 X_{23} \quad e_2 \\ &\vdots \\ y_N &= \alpha + b_1 X_{N_1} + b_2 X_{N_2} + b_3 X_{N_3} + e_N \end{aligned}$$

som gir:

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_N \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & X_{13} \\ 1 & X_{21} & X_{22} & X_{23} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 1 & X_{N_1} & X_{N_2} & X_{N_3} \end{bmatrix} \begin{bmatrix} \alpha \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ e_N \end{bmatrix}$$

som reduseres til den generelle form:

$$\underset{\sim}{y} = X\underset{-}{b} + \underset{-}{e}$$

Vi har her at:

$$X'X = \begin{bmatrix} 1 & 1 & \dots & 1 \\ X_{11} & X_{21} & \dots & X_{N_1} \\ X_{12} & X_{22} & \dots & X_{N_2} \\ X_{13} & X_{23} & \dots & X_{N_3} \end{bmatrix} \begin{bmatrix} 1 & X_{11} & X_{12} & X_{13} \\ 1 & X_{21} & X_{22} & X_{23} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 1 & X_{N_1} & X_{N_2} & X_{N_3} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{N} 1^2 & \sum_{i=1}^{N} 1\cdot X_{i_1} & \sum_{i=1}^{N} 1\cdot X_{i_2} & \sum_{i=1}^{N} 1\cdot X_{i_3} \\ \sum_{i=1}^{N} X_{i_1}\cdot 1 & \sum_{i=1}^{N} X_{i_1}^2 & \sum_{i=1}^{N} X_{i_1}X_{i_2} & \sum_{i=1}^{N} X_{i_1}X_{i_3} \\ \sum_{i=1}^{N} X_{i_2}\cdot 1 & \sum_{i=1}^{N} X_{i_2}X_{i_1} & \sum_{i=1}^{N} X_{i_2}^2 & \sum_{i=1}^{N} X_{i_2}X_{i_3} \\ \sum_{i=1}^{N} X_{i_3}\cdot 1 & \sum_{i=1}^{N} X_{i_3}X_{i_1} & \sum_{i=1}^{N} X_{i_3}X_{i_2} & \sum_{i=1}^{N} X_{i_3}^2 \end{bmatrix}$$

$$= \begin{bmatrix} N & \Sigma X_1 & \Sigma X_2 & \Sigma X_3 \\ \Sigma X_1 & \Sigma X_1^2 & \Sigma X_1 X_2 & \Sigma X_1 X_3 \\ \Sigma X_2 & \Sigma X_1 X_2 & \Sigma X_2^2 & \Sigma X_2 X_3 \\ \Sigma X_3 & \Sigma X_1 X_3 & \Sigma X_2 X_3 & \Sigma X_3^2 \end{bmatrix}$$

siden $\Sigma X_i X_j = \Sigma X_j X_i$

Vi har videre at:

$$X'Y = \begin{bmatrix} 1 & 1 & \dots & 1 \\ X_{11} & X_{21} & \dots & X_{N_1} \\ X_{12} & X_{22} & \dots & X_{N_2} \\ X_{13} & X_{23} & \dots & X_{N_3} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ y_N \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N} 1\cdot y_i \\ \sum_{i=1}^{N} X_{i_1} y_i \\ \sum_{i=1}^{N} X_i\ y_i \\ \sum_{i=1}^{N} X_{i_3} y_i \end{bmatrix} = \begin{bmatrix} \Sigma y \\ \Sigma X_1 y \\ \Sigma X_2 y \\ \Sigma X_3 y \end{bmatrix}$$

C. Enveis-variansanalyse

En enveis-variansanalyse modell kan skrives:

$$y_{ij} = \mu + \alpha_i + e_{ij} \text{ , } i = 1, 2, 3, \ldots , q \text{ , } j = 1, 2, \ldots , n_i$$

Hvis $n_i = p$ for alle i, er materialet balansert dvs. antallet for hver klasse er likt. Når alle observasjonene skrives får en:

$$
\begin{array}{lllllll}
y_{11} & = \mu & + \alpha_1 & & & + e_{11} \\
y_{12} & = \mu & + \alpha_1 & & & + e_{12} \\
\cdot \\
\cdot \\
\cdot \\
y_{1n_1} & = \mu & + \alpha_1 & & & + e_{1n_1} \\
y_{21} & = \mu & & + \alpha_2 & & + e_{21} \\
\cdot \\
\cdot \\
\cdot \\
y_{2n_2} & = \mu & & + \alpha_2 & & + e_{2n_2} \\
\cdot \\
\cdot \\
\cdot \\
y_{q1} & = \mu & & & + \alpha_q & + e_{q1} \\
\cdot \\
\cdot \\
\cdot \\
y_{qn_q} & = \mu & & & + \alpha_q & + e_{qn_q}
\end{array}
$$

som gir:

$$
\begin{bmatrix} y_{11} \\ y_{12} \\ \cdot \\ \cdot \\ \cdot \\ y_{1n_1} \\ y_{21} \\ \cdot \\ \cdot \\ \\ y_{2n_2} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ y_{q1} \\ \cdot \\ \cdot \\ \cdot \\ y_{qn_q} \end{bmatrix}
=
\begin{bmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \\ 1 \\ \cdot \\ \cdot \\ \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix}
\begin{bmatrix} \mu \end{bmatrix}
+
\begin{bmatrix}
1 & 0 & 0 & \cdots & 0 \\
1 & 0 & 0 & \cdots & 0 \\
\cdot & \cdot & \cdot & \cdots & 0 \\
\cdot & \cdot & \cdot & \cdots & \\
\cdot & \cdot & \cdot & \cdots & \\
1 & 0 & 0 & \cdots & 0 \\
0 & 1 & 0 & \cdots & 0 \\
\cdot & \cdot & \cdot & \cdots & \cdot \\
 & & & & \\
 & & & & \\
0 & 1 & 0 & \cdots & 0 \\
\cdot & & & & \\
\cdot & & & & \\
\cdot & & & & \\
\cdot & & & & \\
0 & 0 & 0 & \cdots & 1 \\
\cdot & \cdot & \cdot & \cdots & \cdot \\
\cdot & \cdot & \cdot & \cdots & \cdot \\
\cdot & \cdot & \cdot & \cdots & \cdot \\
0 & 0 & 0 & \cdots & 1
\end{bmatrix}
\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \cdot \\ \cdot \\ \cdot \\ \alpha_q \end{bmatrix}
+
\begin{bmatrix} e_{11} \\ e_{12} \\ \cdot \\ \cdot \\ \cdot \\ e_{1n_1} \\ e_{21} \\ \cdot \\ \cdot \\ \cdot \\ e_{2n_2} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ e_{q1} \\ \cdot \\ \cdot \\ \cdot \\ e_{qn_q} \end{bmatrix}
$$

Dette kan da skrives:

$$\underline{y} = 1\underline{\mu} + X_1\underline{\alpha} + \underline{e}$$

hvor 1 er en $N \times 1$ vektor av entall.

$$\underline{y} = [1 : X_1] \begin{bmatrix} \underline{\mu} \\ \underline{\alpha} \end{bmatrix} + \underline{e}$$

som reduseres til den generelle form:

$$= X\underline{b} + \underline{e}$$

Det er särlig to egenskaper ved X_1 som en bör merke seg. Det for det förste at det eksisterer bare et entall og resten nuller pr. rekke (linje) av X_1. For det andre vil antallet av entall pr. kolonne väre n_i, der i er den i^{te} kolonne, $i = 1, 2, \dots, q$. En matrise med disse egenskaper går ofte under navnet "incidence" matrise.

$$X'X = \{1 : X_1\}' \{1 : X_1\}$$

$$= \begin{bmatrix} 1' \\ X_1' \end{bmatrix} [1 : X_1] = \begin{bmatrix} N & 1'X_1 \\ X_1'1 & X_1'X_1 \end{bmatrix}$$

$$1'X_1 = \{n_1 \quad n_2 \quad n_3 \; . \; . \; . \; n_q\}$$

$$X_1'1 = \{1'X_1\}'$$

$$X_1'X_1 = \begin{bmatrix} n_1 & 0 & 0 & \dots & 0 \\ 0 & n_2 & 0 & \dots & 0 \\ 0 & 0 & n_3 & \dots & 0 \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ 0 & 0 & & \dots & n_q \end{bmatrix}$$

Derfor er:

$$X'X = \begin{bmatrix} N & n_1 & n_2 & \dots & n_q \\ n_1 & n_1 & 0 & \dots & 0 \\ n_2 & 0 & \cdot & & \cdot \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & & \cdot \\ n_q & 0 & & \dots & n_q \end{bmatrix}$$

$$X'Y = \{1 : X_1\}' \, Y$$

$$= \begin{bmatrix} 1' \\ X_1' \end{bmatrix} Y = \begin{bmatrix} y_{\cdot\cdot} \\ X_1'Y \end{bmatrix}$$

$$= \begin{bmatrix} y_{\cdot\cdot} \\ \Sigma y_{1j} \\ \Sigma y_{2j} \\ \cdot \\ \cdot \\ \cdot \\ \Sigma y_{qj} \end{bmatrix} = \begin{bmatrix} y_{\cdot\cdot} \\ y_{1\cdot} \\ y_{2\cdot} \\ \cdot \\ \cdot \\ \cdot \\ y_{q\cdot} \end{bmatrix}$$

D. Toveis-variansanalyse

En toveis-variansanalysemodell kan skrives:

$$y_{ijk} = \mu + a_i + b_j + e_{ijk}$$

hvor $i = 1, 2, 3, \ldots, q$, $j = 1, 2, 3, \ldots, p$, og n_{ij} er antall observasjoner i den ij^{te} cella. Denne modellen kan skrives:

$$\underline{y} = 1\underline{\mu} + X_1\underline{a} + X_2\underline{b} + \underline{e}$$

hvor $\underline{y}$ er en $N \times 1$ vektor av observasjoner

X_1 og X_2 er "incidence" matriser med dimensjonen $N \times q$ og $N \times p$ henholdsvis. I den generelle form gir dette:

$$\underline{y} = [1 : X_1 : X_2] \begin{bmatrix} \mu \\ a \\ b \end{bmatrix} + \underline{e}$$

$$= X\underline{b} + \underline{e}$$

Vi har så at:

$$X'X = [1 : X_1 : X_2]' \, [1 : X_1 : X_2]$$

$$= \begin{bmatrix} 1' \\ X_1' \\ X_2' \end{bmatrix} \quad [1 : X_1 : X_2]$$

$$= \begin{bmatrix} 1'1 & 1'X_1 & 1'X_2 \\ X_1'1 & X_1'X_1 & X_1'X_2 \\ X_2'1 & X_2'X_1 & X_2'X_2 \end{bmatrix}$$

hvor $1'1 \;=\; N \;=\; \sum_{i=1}^{q} n_{i.} \;=\; \sum_{j=1}^{p} n_{.j}$

$$1'X_1 = [n_{1.} \quad n_{2.} \quad . \; . \; . \quad n_{q.}]$$

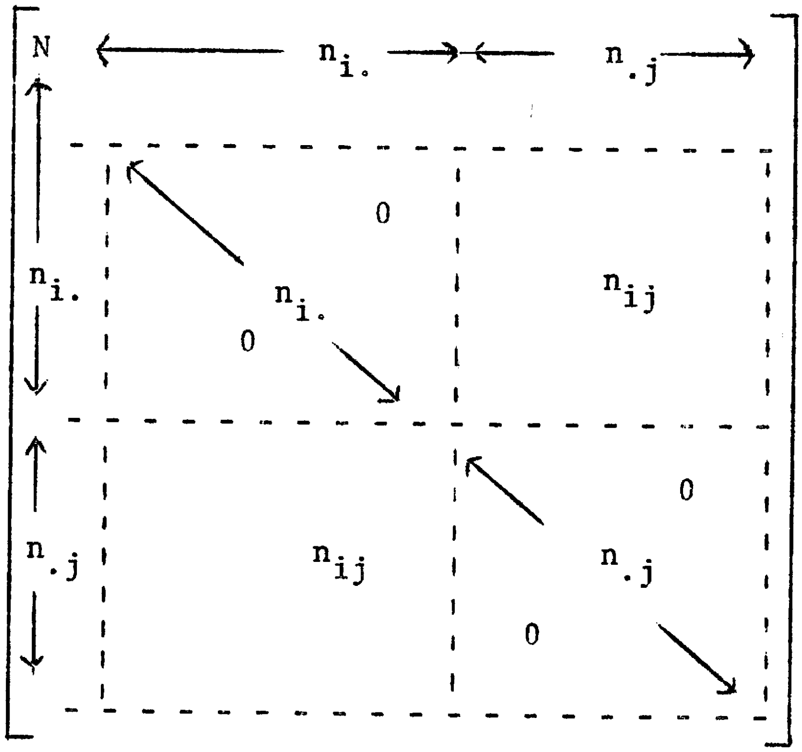
$$1'X_2 = [n_{.1} \quad n_{.2} \quad . \; . \; . \quad n_{.p}]$$

$$X_1'X_1 = \begin{bmatrix} n_{1.} & 0 & 0 \; . \; . \; . & 0 \\ 0 & n_{2.} & . \; . \; . \; . & . \\ . & & & . \\ . & & & . \\ 0 & . & . & n_{q.} \end{bmatrix}$$

$$X_1'X_2 = \begin{bmatrix} n_{11} & n_{12} & . \; . \; . & n_{1p} \\ n_{21} & n_{22} & . \; . \; . & n_{2p} \\ . & & & . \\ . & & & . \\ n_{q1} & & & n_{qp} \end{bmatrix}$$

$$X_2'X_2 = \begin{bmatrix} n_{.1} & 0 & . \; . \; . & 0 \\ 0 & n_{.2} & & 0 \\ . & & . & . \\ . & & \;\; . & . \\ . & & \;\;\;\; . & . \\ 0 & . & . \; . & n_{.p} \end{bmatrix}$$

så X'X vil bli av typen:



Den viktigste konklusjon fra disse matrise-uttrykkene er at X'X er alltid symmetrisk. Generelt gjelder det at for en gitt matrise X vil X'X og XX' väre symmetrisk.

En viktig egenskap ved en matrise er antallet av lineäre uavhengige kolonner eller rekker (linjer). Denne egenskap er definert som rank og vil bli symbolisert ved r (X) (Searle, 1967;"Chapter 5)".

For en matrise X'X med dimensjonen $k \times k$ vil det bare eksistere en invers av X'X, hvis r(X'X) = k. Matriser av typen X'X vil som regel ha en rank som er lik antall frihetsgrader. Derfor vil regresjonsanalyse medföre at X'X er full rank, mens variansanalysemodeller ("design models") förer til at rank er mindre enn dimensjonen av matrisen X'X.

Hvis en matrise av dimensjonen $N \times k$ har rank som er mindre enn k, vil en eller flere av kolonnene kunne uttrykkes ved hjelp av de uavhengige kolonner. En matrise gitt ved:

$$X = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

har r $(X) = 2$. La oss definere $X = \{X_1 : X_2\}$ hvor X_1 er 4×2 og r $(X_1) = 2$. En kan da finne at:

$$X_2 = X_1 C$$

dvs. X_2 er en lineär kombinasjon av X_1

$$X_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

så $C = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ tillfredsstiller denne likhet.

For statistiske modeller vil en ofte si at X av dimensjonen $N \times k$ ikke er av full rank hvis r $(X) < k$. En kan derfor uttrykke den delen av en matrise som ikke er av full rank (X_2) ved hjelp av den delen som er av full rank (X_1) for alle typer av matriser. Dette vil bli närmere behandlet senere.

E. Lösning til normal-ligningene

La oss som tidligere forutsette at en random variabel kan beskrives ved fölgende lineäre modell:

$$\underline{y} = X\underline{b} + \underline{e}$$

hvor $\underline{y}$ er en $N \times 1$ vektor av en observerbar random variabel

X er en $N \times k$ matrise av konstanter

$\underline{b}$ er en $k \times 1$ vektor av ukjente parametre som skal estimeres

$\underline{e}$ er en $N \times 1$ vektor av en ikke-observerbar random variabel

Den minste mengde (set) av forutsetninger som må väre tilfredsstilt er:

i) $E(\underline{e}) = 0$

ii) $E(\underline{e}\underline{e}') = I\sigma_e^2$, hvor I er en $N \times N$ enhetsmatrise (identity)

iii) X er en gitt mengde av konstanter

Som en konsekvens av den mengde forutsetninger som er definert ved i, ii og iii, kan fölgende bli avledet:

iv) $E(\underline{y}) = E(X\underline{b} + \underline{e}) = X\underline{b}$ fra i og iii

v) $E(\underline{b}\underline{e}') = \underline{b}E(\underline{e}') = \underline{b}0 = 0$ fra i

Den förste forutsetning sier at $E(e_i) = 0$ for alle i dvs. at e_i er en random variabel hvor forventningen er lik null. Forutsetning (ii) kan en vise på fölgende måte:

$$E(\underline{e}\underline{e}') = E \begin{bmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ \cdot \\ e_N \end{bmatrix} [e_1 \quad e_2 \quad \ldots \quad e_N] =$$

$$= \begin{bmatrix} E(e_1^2) & E(e_1,e_2) & \ldots & E(e_1,e_N) \\ E(e_2,e_1) & E(e_2^2) & \ldots & E(e_2,e_N) \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ E(e_N,e_1) & E(e_N,e_2) & \ldots & E(e_N^2) \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_e^2 & 0 & 0 \\ 0 & \sigma_e^2 & 0 \\ \cdot & \cdot & \cdot \\ 0 & 0 & \sigma_e^2 \end{bmatrix} = I\sigma_e^2$$

Uttrykkene på hoved-diagonalen ("the main diagonal") viser at $E(e_i^2) = \sigma_e^2$ for alle i dvs. e_i har konstant varians som er lik σ_e^2. Elementene utenom hoved-diagonalen ("the off-diagonals") viser også at $E(e_t, e_{t+s}) = 0$ for $s \neq 0$, dvs. e_i er parvis ukorrelerte (hvis $e_i \sim N(0, \sigma_e^2)$ betyr dette at e_i er parvis uavhengige, tegnet ~ erstatter "er fordelt som"). Forutsetning (iii) er slik å forstå at X er en fiksert mengde ("set") av verdier dvs. X er ikke framkommet som et resultat av en sampling prosess (Graybill, 1961 s. 109). Hvis X er en uavhengig regresjonsvariabel med verdier mellom 50 og 100, vil en ny sampling av Y som gir X verdier mellom 100 og 150 ikke nödvendigvis gi samme regresjons-funksjon. Derfor vil egenskapene ("the properties") av de utviklete estimatorer og deres tester väre totalt betinget av den gitte X.

Siden $\underline{b}$ er en vektor av konstanter, og X er en fiksert matrise, vil vektoren $\underline{e}$ tilfredsstille en random vektor bare i de til-feller at $\underline{y}$ er en random vektor. Dette kan vi vise ved fölgende ligning:

$$\underline{e} = \underline{y} - X\underline{b}$$

Derfor er den ikke-observerbare $\underline{e}$ en lineär funksjon av den observerbare $\underline{y}$ og er dermed en random variabel hvis $\underline{y}$ tilfreds-stiller dette kravet i tillegg til forutsetning (iii).

Med basis fra den observerte matrise X og den observerte random variabel $\underline{y}$, skal vi nå utvikle estimatorer for $\underline{b}$ og σ_e^2 og under-söke en del egenskaper ved disse estimatorer.

Ved hjelp av ligningen $\underline{y} = X\underline{b} + \underline{e}$ kan <u>restkvadratsummen</u> for den random variable bli utledet:

$$\begin{aligned}\Sigma e_i^2 &= \underline{e}'\underline{e} = (\underline{y} - X\underline{b})'\ (\underline{y} - X\underline{b}) \\ &= (\underline{y}' - \underline{b}'X')(\underline{y} - X\underline{b}) = \underline{y}'\underline{y} - \underline{b}'X'\underline{y} - \underline{y}'X\underline{b} + \underline{b}'X'X\underline{b} \\ &= \underline{y}'\underline{y} - 2\underline{b}'X'\underline{y} + \underline{b}'X'X\underline{b}\end{aligned}$$

siden $\underline{y}'X\underline{b}$ er en "scalar" og derfor $\underline{y}'X\underline{b} = (\underline{y}'X\underline{b})' = \underline{b}'X'\underline{y}$

For å finne den vektor av $\underline{b}$ som gir minimum kvadratsumrest, differensierer vi ligningen ovenfor med hensyn på $\underline{b}$:

$$\frac{\delta\ (\underline{e}'\underline{e})}{\delta\ \underline{b}} = -\ 2X'\underline{y} + 2X'X\underline{b}$$

Ved å sette denne ligningen lik null får en fölgende:

$$X'X\underline{b} = X'\underline{y}$$

som kalles normal-ligningene. Fra normal-ligningene ser en at $X'X\underline{b}$ er en lineär funksjon av den random variable $\underline{y}$. Det er derfor avgjörende at $\underline{y}$ er en vektor av representative verdier for den populasjon som en vil applisere resultatene på, dvs. det må väre en random variabel.

Om e-elementene er korrelerte, kan en på samme sett bevise at normal-ligningene er:

$$XV^{-1}X\underline{b} = X'V^{-1}\underline{y}$$

hvor V er var.-kov.-matrisen for $\underline{e}$ [dvs. $e'e = (\underline{y} - X\underline{b})'V^{-1}(\underline{y} - X\underline{b})$].

III. METODE TIL Å BESTEMME HVA SOM ER ESTIMÉRBART OG IKKE ESTIMÉRBART I STATISTISKE MODELLER

Til grunn for en vektor av observasjoner ligger det en statistisk modell. Denne modellen kan inneholde flere effekter (variasjonsårsaker), og den enkelte effekt kan väre "fixed" eller "random". For selve estimeringen må en avgjöre hva som er estimérbart og ikke estimérbart i modellen.

La den generelle modellen väre som vist nedenfor:

$$\underline{y} = X\underline{b} + \underline{e}$$

$\underline{y}$ = vektor av observasjoner

$\underline{b}$ = vektor som inneholder "fixed" effekter

X = matrise ("designed matrix") som spesifiserer $\underline{b}$

$\underline{e}$ = vektor som inneholder "random" elementer

Det enkleste tilfellet for estimering har en når $\underline{X}$ er av full rank (r = rank). Med full rank forstår en at ikke noe avhengighetsforhold eksisterer mellom rekker eller mellom kolonner.

Hvis $\underline{X}$ ikke er av full rank blir spörsmålet om estimering mer komplisert. I det fölgende vil en gå inn på metoder til å bestemme hva som er estimérbart i slike tilfeller. En vil forelöpig bare påpeke at r(X'X) betyr rank av X'X.

A. Samspill er tilstede

La oss betrakte fölgende numeriske eksempel:

Frekvenstabell

	b_1	b_2	b_3	b_4	Sum
a_1	2	3	1	2	8
a_2	1	-	2	1	4
a_3	2	-	3	-	5
Sum	5	3	6	3	17

De tilhörende summer er:

	b_1	b_2	b_3	b_4	Sum
a_1	8	15	2	6	31
a_2	3	-	6	2	11
a_3	12	-	6	-	18
Sum	23	15	14	8	60

Modellen for dette eksemplet kan i detalj sies å väre:

$$Y_{ijk} = \mu + a_i + b_j + ab_{ij} + e_{ijk}\text{ , hvor}$$

μ ; a_i $(i = 1, 2, 3)$; b_j $(j = 1, 2, 3, 4)$ er "fixed".

e_{ijk} er "random" med $V(e) = I\sigma^2$ (V = varians).

Uttrykt ved hjelp av den förstnevnte modellen ($\underline{y} = X\underline{b} + \underline{e}$) får en:

$$\underline{b}' = (\mu\ a_1\ a_2\ a_3\ b_1\ b_2\ b_3\ b_4\ ab_{11}\ ab_{12}\ ab_{13}\ ab_{14}\ ab_{21}\ ab_{23}\ ab_{24}\ ab_{31}\ ab_{33})$$

$$\underline{y} = \begin{bmatrix} Y_{111} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ Y_{333} \end{bmatrix}$$

X-matrisen er satt opp på side III 3.

En er interessert i å finne ranken av $X'X$ (en forutsetter at $X'V^{-1}X = X'X$). La oss nå sette opp $X'X\hat{\underline{b}} = X'\underline{y}$ numerisk. En henviser til side III 4.

Ved å studere $X'X\hat{\underline{b}}$ kan en se fölgende:

$\hat{\mu}$-ligningen = $\sum\limits_i \hat{a}_i$-ligningene = $\sum\limits_j \hat{b}_j$-ligningene = $\sum\limits_i \sum\limits_j \hat{ab}_{ij}$-ligningene.

$$X = \begin{array}{c}
\begin{array}{cccccccc|ccccccccc}
\mu & a_1 & a_2 & a_3 & b_1 & b_2 & b_3 & b_4 & ab_{11} & ab_{12} & ab_{13} & ab_{14} & ab_{21} & ab_{23} & ab_{24} & ab_{31} & ab_{33}
\end{array} \\
\left[\begin{array}{cccccccc|ccccccccc}
1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{array}\right]
\end{array}$$

$$
\overset{X'X}{\left[\begin{array}{c|ccc|cccc|cccc|ccc|cc}
17 & 8 & 4 & 5 & 5 & 3 & 6 & 3 & 2 & 3 & 1 & 2 & 1 & 2 & 1 & 2 & 3 \\
8 & 8 & 0 & 0 & 2 & 3 & 1 & 2 & 2 & 3 & 1 & 2 & 0 & 0 & 0 & 0 & 0 \\
4 & 0 & 4 & 0 & 1 & 0 & 2 & 1 & 0 & 0 & 0 & 0 & 1 & 2 & 1 & 0 & 0 \\
5 & 0 & 0 & 5 & 2 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 3 \\
\hline
5 & 2 & 1 & 2 & 5 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 1 & 0 & 0 & 2 & 0 \\
3 & 3 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
6 & 1 & 2 & 3 & 0 & 0 & 6 & 0 & 0 & 0 & 1 & 0 & 0 & 2 & 0 & 0 & 3 \\
3 & 2 & 1 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 2 & 0 & 0 & 1 & 0 & 0 \\
\hline
2 & 2 & 0 & 0 & 2 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
3 & 3 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
2 & 2 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\
\hline
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
2 & 0 & 2 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
\hline
2 & 0 & 0 & 2 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \\
3 & 0 & 0 & 3 & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3
\end{array}\right]}
\overset{\hat{\underline{b}}}{\begin{bmatrix}
\hat{\mu} \\ \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{b}_1 \\ \hat{b}_2 \\ \hat{b}_3 \\ \hat{b}_4 \\ \hat{ab}_{11} \\ \hat{ab}_{12} \\ \hat{ab}_{13} \\ \hat{ab}_{14} \\ \hat{ab}_{21} \\ \hat{ab}_{23} \\ \hat{ab}_{24} \\ \hat{ab}_{31} \\ \hat{ab}_{33}
\end{bmatrix}}
=
\overset{X'\underline{y}}{\begin{bmatrix}
60 \\ 31 \\ 11 \\ 18 \\ 23 \\ 15 \\ 14 \\ 8 \\ 8 \\ 15 \\ 2 \\ 6 \\ 3 \\ 6 \\ 2 \\ 12 \\ 6
\end{bmatrix}}
$$

Full rank

Videre, hver $\hat{a}_i$-ligning = $\sum_j \hat{ab}_{ij}$-ligningene

hver $\hat{b}_j$-ligning = $\sum_i \hat{ab}_{ij}$-ligningene

Ut fra dette eller fra mer konvensjonelle metoder til å bestemme rank (Searle, 1967) kan en konkludere at $r(X'X) = 9$, dvs. antall celler med observasjoner i hver av tabellene på side III 1 og 2. En submatrise som er av full rank i X'X er markert på side III 4.

<u>Hvilke ledd eller funksjoner er så estimérbare?</u> La oss sette $X = (X_1 X_2)$ hvor X_1 er de 9 <u>siste</u> kolonnene (seksjonen av full rank) av X og $X_1 C = X_2$ dvs. de förste 8 kolonnene av X (seksjonen av ikke full rank). Dette betyr at $\underline{C}$ har fölgende form:

$$C_{9x8} = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Videre la vektoren λ väre av samme dimensjon som $\underline{b}$. De funksjonene en önsker å estimere kan da sies å väre av fölgende form:

$$\lambda'\underline{b} = [\underbrace{\mu^* a_1^* \; .... \; b_4^*}_{\alpha_2'} \quad \underbrace{ab_{11}^* \; .... \; ab_{33}^*}_{\alpha_1'}]\,\underline{b}$$ hvor koeffisientene

μ^*, a_1^* osv. kan väre 0 eller en konstant forskjellig fra 0. En kan f.eks. nevne at hvis en önsker å estimere μ, blir:

$$\lambda'\underline{b} = [1 \quad 0 \; .... \; 0] \begin{bmatrix} \mu \\ \vdots \\ ab_{33} \end{bmatrix} = \mu \quad \text{hvor} \quad \mu^* = 1.$$

For å finne de restriksjonene ("constraints") som en må ta hensyn til ved vurdering av estimérbarhet foretar en fölgende multiplikasjon:

$$\alpha_1' C = [ab_{11}^* \quad ab_{12}^* \quad ab_{13}^* \quad ab_{14}^* \quad ab_{21}^* \quad ab_{23}^* \quad ab_{24}^* \quad ab_{31}^* \quad ab_{33}^*]$$

$$
\underbrace{\begin{bmatrix}
1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 1 & 0
\end{bmatrix}}_{C} =
$$

$$
\begin{bmatrix}
ab^*_{11} + ab^*_{12} + ab^*_{13} + ab^*_{14} + ab^*_{21} + ab^*_{23} + ab^*_{24} + ab^*_{31} + ab^*_{33} \\
ab^*_{11} + ab^*_{12} + ab^*_{13} + ab^*_{14} \\
ab^*_{21} + ab^*_{23} + ab^*_{24} \\
ab^*_{31} + ab^*_{33} \\
ab^*_{11} + ab^*_{21} + ab^*_{31} \\
ab^*_{12} + \\
ab^*_{13} + ab^*_{23} + ab^*_{33} \\
ab^*_{14} + ab^*_{24}
\end{bmatrix}
$$

De betingelsene som må väre oppfylte når en vurderer estimérbarhet er derfor at (NB på venstresiden av ligningene har vi koeffisientene i α_2' ettersom $X_1 C = X_2$):

(1) $\mu^* = ab^*_{11} + ab^*_{12} + ab^*_{13} + ab^*_{14} + ab^*_{21} + ab^*_{23} + ab^*_{24} + ab^*_{31} + ab^*_{33}$

(2) $a^*_1 = ab^*_{11} + ab^*_{12} + ab^*_{13} + ab^*_{14}$

(3) $a^*_2 = ab^*_{21} + ab^*_{23} + ab^*_{24}$

(4) $a^*_3 = ab^*_{31} + ab^*_{33}$

(5) $b^*_1 = ab^*_{11} + ab^*_{21} + ab^*_{31}$

(6) $b^*_2 = ab^*_{12}$

(7) $b_3^* = ab_{13}^* + ab_{23}^* + ab_{33}^*$

(8) $b_4^* = ab_{14}^* + ab_{24}^*$

La oss ta noen eksempler for å illustrere metoden:

a. Er μ estimérbar? Ifölge betingelsen ovenfor må koeffisienten for μ (altså μ^*) väre lik summen av koeffisientene for samspill ($ab_{11}^* + \ldots + ab_{33}^*$; se (1)). Når vi bare betrakter μ, vil koeffisienten for μ väre 1, mens koeffisientene for samspill (ab_{ij}^*) er null. [$\lambda'\underline{b} = \mu$ betyr $\lambda' = (1 \quad 0 \quad 0 \ldots 0)$]. Altså betingelsen i (1) er ikke oppfylt [$1 \neq 0 + \ldots + 0$], og derfor er ikke μ estimérbar.

b. Er $\mu + a_1 + b_1 + ab_{11}$ estimérbar? Tar vi först for oss μ, ser en at koeffisienten for μ er lik koeffisienten for ab_{11} (se (1)). Videre er koeffisienten for a_1 (se (2)) og b_1 (se (5)) lik koeffisienten for ab_{11}. En kan derfor konkludere at $\mu + a_1 + b_1 + ab_{11}$ er estimérbar.

c. Er $\mu + a_1$ estimérbar? Ifölge (1) og (2) er summen av koeffisientene på höyresiden av likhetstegnet lik null. Koeffisientene på venstresiden er (1) . Fölgelig er $\mu + a_1$ ikke estimérbar.

d. Er $a_1 - a_2$ estimérbar? Ved å sjekke (2) og (3) kan en slutte at denne differansen ikke er estimérbar (koeffisientene for samspill er null).

e. Er $\mu + a_1 + \bar{b}. + \overline{ab}_{1.}$ estimérbar? I dette tilfellet vil rekkevektoren foran kolonnevektoren $\hat{\underline{b}}$ bli som fölger:

$$(1 \quad 1 \quad 0 \quad 0 \quad \tfrac{1}{4} \quad \tfrac{1}{4} \quad \tfrac{1}{4} \quad \tfrac{1}{4} \quad \tfrac{1}{4} \quad \tfrac{1}{4} \quad \tfrac{1}{4} \quad \tfrac{1}{4} \quad 0 \quad 0 \quad 0 \quad 0 \quad 0)$$

De aktuelle betingelser er oppfylte, og funksjonen ovenfor er estimérbar.

B. Samspill er ikke tilstede

Fra side III 4 ser en at $X'X\hat{\underline{b}}$ vil ta denne formen når samspill-leddene utelates:

$$\left[\begin{array}{c|ccc|cccc}
17 & 8 & 4 & 5 & 5 & 3 & 6 & 3 \\
\hline
8 & 8 & 0 & 0 & 2 & 3 & 1 & 2 \\
4 & 0 & 4 & 0 & 1 & 0 & 2 & 1 \\
5 & 0 & 0 & 5 & 2 & 0 & 3 & 0 \\
\hline
5 & 2 & 1 & 2 & 5 & 0 & 0 & 0 \\
3 & 3 & 0 & 0 & 0 & 3 & 0 & 0 \\
6 & 1 & 2 & 3 & 0 & 0 & 6 & 0 \\
3 & 2 & 1 & 0 & 0 & 0 & 0 & 3
\end{array}\right]
\begin{bmatrix} \hat{\mu} \\ \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{b}_1 \\ \hat{b}_2 \\ \hat{b}_3 \\ \hat{b}_4 \end{bmatrix}$$

Her er $\hat{\mu}$ ligningen = $\sum_i \hat{a}_i$ ligningene = $\sum_j \hat{b}_j$ ligningene, og $r(X'X) = 6$. For å danne en matrise ("designed") av full rank flytter en så venstre kolonne av $\underline{X}$ over til höyre, således:

$$\begin{array}{cc}
\begin{array}{cccccc} a_1 & a_2 & a_3 & b_1 & b_2 & b_3 \end{array} & \begin{array}{cc} b_4 & \mu \end{array} \\
\underbrace{\begin{bmatrix}
1 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 & 0 & 1
\end{bmatrix}}_{X_1 \text{ (full rank)}} &
\underbrace{\begin{bmatrix}
0 & 1 \\
0 & 1 \\
0 & 1 \\
0 & 1 \\
0 & 1 \\
0 & 1 \\
1 & 1 \\
1 & 1 \\
0 & 1 \\
0 & 1 \\
0 & 1 \\
1 & 1 \\
0 & 1 \\
0 & 1 \\
0 & 1 \\
0 & 1 \\
0 & 1
\end{bmatrix}}_{X_2}
\end{array}$$

Det neste blir da å finne $\underset{=}{C}$. Denne kan finnes ut fra fölgende identitet:

$X_1 C = X_2$, og derfor er

$$C = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ -1 & 0 \\ -1 & 0 \\ -1 & 0 \end{bmatrix}$$

Fölgelig

$$\alpha_1' C = [a_1^* \quad a_2^* \quad a_3^* \quad b_1^* \quad b_2^* \quad b_3^*] \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ -1 & 0 \\ -1 & 0 \\ -1 & 0 \end{bmatrix} =$$

$$= \begin{bmatrix} a_1^* + a_2^* + a_3^* - b_1^* - b_2^* - b_3^* \\ a_1^* + a_2^* + a_3^* \end{bmatrix}$$

Videre er $\alpha_2' = [b_4^* \ \mu^*]$

Betingelsene for estimérbarhet blir da:

(1) $b_4^* = a_1^* + a_2^* + a_3^* - b_1^* - b_2^* - b_3^*$

(2) $\mu^* = a_1^* + a_2^* + a_3^*$

La oss ta noen eksempler i dette tilfellet også:

a. Er μ estimérbar? I så fall må koeffisienten for μ^* väre lik summen $a_1^* + a_2^* + a_3^*$ (se (2)). Denne betingelsen er ikke oppfylt, og μ er ikke estimérbar.

b. Er a_1-a_2 estimérbar? Her skal $b_4^* = 0$ hvilket den er (se (1)). En kan også merke seg at a_1-a_2 ikke var estimérbar når samspill var inkludert.

c. Er $\mu + a_1 + \sum_j \lambda_j b_j$ estimérbar? La oss anta at $\sum_j \lambda_j = 1$.

En får da fölgende rekkevektor foran $\hat{\underline{b}}$:

$$\begin{bmatrix} 1 & 1 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}$$

Koeffisienten for μ er lik $a_1^* + a_2^* + a_3^*$. Det foreligger heller ikke noen uoverstemmelse med 1. Derfor er $\mu + a_1 + \sum_j \lambda_j b_j$ estimérbar. Det samme kan sies om $\mu + \sum_i \lambda_i a_i + b_j$ hvor $\sum_i \lambda_i = 1$.

En kan også merke seg at når <u>samspill ikke eksisterer så kan en estimere gjennomsnittet for en celle (subclass) hvor $n_{ij} = 0$. Estimatet er $\hat{\mu} + \hat{a}_i + \hat{b}_j$.</u>

C. Enklere metode til å bestemme estimérbarhet

Den metoden som er beskrevet i det foregående kan sies å väre den sikreste for å finne hvilke ledd eller funksjoner som er estimérbare. Ulempen er imidlertid at den kan väre noe komplisert. I det fölgende vil en gå inn på en noe forenklet metode som först og fremst kan väre berettiget når samspill er tilstede.

La koeffisientene for ab_{ij} bli betegnet med λ_{ij} $(n_{ij} > 0)$, koeffisienten for a_i med $\lambda_{i.}$ og koeffisienten for b_j med $\lambda_{.j}$. Koeffisienten for μ blir da $\lambda_{..}$. De estimérbare funksjonene vil da väre av fölgende form:

$\sum_i \sum_j \lambda_{ij}(\mu + a_i + b_j + ab_{ij})$, hvor alle $\lambda_{ij} = 0$ for $n_{ij} = 0$. (n_{ij} er frekvensen av ij^{te} undergruppe).

En setter så opp en tabell hvor "fixed" effektene og de tilhörende koeffisienter inngår:

	b_1	b_2	b_3	b_4	
a_1	λ_{11}	λ_{12}	λ_{13}	λ_{14}	$\lambda_{1.}$
a_2	λ_{21}	-	λ_{23}	λ_{24}	$\lambda_{2.}$
a_3	λ_{31}	-	λ_{33}	-	$\lambda_{3.}$
	$\lambda_{.1}$	$\lambda_{.2}$	$\lambda_{.3}$	$\lambda_{.4}$	$\lambda_{..}$

Den videre fremgangsmåten vil så bli illustrert gjennom en del eksempler.

a. Er a_1 estimérbar? I dette tilfellet blir tabellen som nedenfor:

	b_1	b_2	b_3	b_4	
a_1	0	0	0	0	1
a_2	0	-	0	0	0
a_3	0	-	0	-	0
	0	0	0	0	0

Koeffisienten for $\hat{\mu}$

En ser av denne tabellen at blant annet $\lambda_{1.} + \lambda_{2.} + \lambda_{3.} \neq \lambda_{..}$, med andre ord inkonsekvens. Derfor er a_1 ikke estimérbar.

b. Er $\mu + a_1$ estimérbar? En setter så opp den tilhörende tabellen:

	b_1	b_2	b_3	b_4	
a_1	0	0	0	0	1
a_2	0	-	0	0	0
a_3	0	-	0	-	0
	0	0	0	0	1

Her ser en blant annet at $\lambda_{.1} + \lambda_{.2} + \lambda_{.3} + \lambda_{.4} \neq \lambda_{..}$, så $\mu + a_1$ er ikke estimérbar.

c. Er $b_1 - b_2$ estimérbar? Tabellen er da:

	b_1	b_2	b_3	b_4	
a_1	0	0	0	0	0
a_2	0	-	0	0	0
a_3	0	-	0	0	0
	1	-1	0	0	0

En kan konkludere at $b_1 - b_2$ ikke er estimérbar fordi blant annet $\lambda_{11} + \lambda_{21} + \lambda_{31} \neq \lambda_{.1}$

Ut fra disse eksempler skulle det gå frem at enhver funksjon uten ab_{ij} ikke er estimérbar.

d. Er $\mu + a_1 + b_1 + ab_{11}$ estimérbar? Tabellen blir nå:

	b_1	b_2	b_3	b_4	
a_1	1	0	0	0	1
a_2	0	-	0	0	0
a_3	0	-	0	-	0
	1	0	0	0	1

I denne tabellen finner en ingen inkonsekvenser fordi:

$\lambda_{11} + \lambda_{21} + \lambda_{31} = \lambda_{\cdot 1}$

$\lambda_{11} + \lambda_{12} + \lambda_{13} + \lambda_{14} = \lambda_{1\cdot}$

$\lambda_{\cdot 1} + \lambda_{\cdot 2} + \lambda_{\cdot 3} + \lambda_{\cdot 4} = \lambda_{\cdot\cdot}$

$\lambda_{1\cdot} + \lambda_{2\cdot} + \lambda_{3\cdot} = \lambda_{\cdot\cdot}$

Funksjonen er derfor estimérbar.

e. Er $a_1 + \frac{1}{3}(b_1 + ab_{11} + b_3 + ab_{13} + b_4 + ab_{14}) -$

$\left[a_2 + \frac{1}{3}(b_1 + ab_{21} + b_3 + ab_{23} + b_4 + ab_{24})\right]$ estimérbar?

Denne differansen er lik $(a_1 + \overline{ab}_{1\cdot}) - (a_2 + \overline{ab}_{2\cdot})$ hvor kolonnene 1, 3 og 4 inngår i gjennomsnittene $(\overline{ab}_{1\cdot}; \overline{ab}_{2\cdot})$. Tabellen blir nå:

	b_1	b_2	b_3	b_4	
a_1	$\frac{1}{3}$	0	$\frac{1}{3}$	$\frac{1}{3}$	1
a_2	$-\frac{1}{3}$	-	$-\frac{1}{3}$	$-\frac{1}{3}$	-1
a_3	0	-	-	0	0
	$\frac{1}{3}-\frac{1}{3}=0$	0	$\frac{1}{3}-\frac{1}{3}=0$	$\frac{1}{3}-\frac{1}{3}=0$	0

Det foreligger ikke noen inkonsekvenser så differansen er estimérbar.

Fra disse eksemplene kan en legge merke til at ikke noe tilfredsstillende estimat av rekke- eller kolonnemiddel, eller rekke- eller kolonnedifferanser er oppnåelig når enkelte celler er uten observasjoner.

Hva ville situasjonen bli om alle celler hadde observasjoner?

En kan da definere rekkemiddel som $\mu + a_i + \sum_{j=1}^{q} \lambda_j (b_j + ab_{ij})$ hvor $\sum_j \lambda_j = 1$. q er antall fylte celler i den enkelte rekke. I dette eksemplet vil $q = 4$. Hvis $i = 1$, blir tabellen som vist nedenfor:

	b_1	b_2	b_3	b_4	
a_1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	1
a_2	0	0	0	0	0
a_3	0	0	0	0	0
	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	1

Konklusjon: $\mu + a_1 + \sum_{j=1}^{4} \frac{1}{4} (b_j + ab_{1j})$ er estimérbar fordi ingen inkonsekvenser foreligger i tabellen.

Til slutt skal en bare minne om at ved å telle opp antall frihetsgrader vil en oppnå rank for den matrisen (X'X) en arbeider med. En vil illustrere dette for eksemplet på side III 1.

Samspill er tilstede:

μ	tar 1 frihetsgrad
a_i	tar 2 frihetsgrader [3 - 1]
b_j	tar 3 frihetsgrader [4 - 1]
ab_{ij}	tar 3 frihetsgrader [(4 - 1)(3 - 1)-antall tomme celler]
Tilsammen (rang):	9 frihetsgrader

Samspill er ikke tilstede:

μ	tar 1 frihetsgrad
a_i	tar 2 frihetsgrader [3 - 1]
b_j	tar 3 frihetsgrader [4 - 1]
Tilsammen (rang):	6 frihetsgrader

Som en ser blir tallverdiene for rank de samme som tidligere.

I kapittel V vil en også komme noe inn på estimérbarhet.

D. Reparameterisering

En mye benyttet metode for å löse normal-ligningene er reparameterisering, som betyr at for modellen $\underline{y} = X\underline{b} + \underline{e}$ vil vektoren $\underline{b}$ bli transformert til en vektor $\underline{\alpha}$ som er $\underline{\alpha} = U_1\underline{b}$, hvor hvert element av $\underline{\alpha}$ er estimerbart (Graybill, 1961 s. 227-241). Det er spesielt en metode som er benyttet, nemlig summen av effektene for hver klassifisering er lik null.

La oss först bruke fölgende modell for å illustrere metoden (Fimland, 1972):

$$Y_{ij} = \mu + a_i + e_{ij}$$

som kan skrives:

$$\underline{y} = 1\underline{\mu} + Z\underline{a} + \underline{e}$$

hvor 1 er en $N \times 1$ vektor av ett-tall.

Z er en $N \times q$ matrise av en og null, hvor det forekommer bare ett entall og resten null i hver rekke av Z. Definer også $Z = [Z_1 : Z_2]$ hvor Z_2 er $N \times 1$ vektor.

Hvis "constraint" er: $\sum_{i=1}^{q} a_i = 0$ vil $a_q = - \sum_{i=1}^{q-1} a_i$.

Derfor kan vi erstatte a_q med $- \sum_{i=1}^{q-1} a_i$ for de observasjoner hvor a_q opptrer, samtidig må modellen skrives:

$$Y_{ij} = \mu + a_i + \frac{\sum_{i=1}^{q} a_i}{q} - \frac{\sum_{i=1}^{q} a_i}{q} + e_{ij}$$

$$= \underbrace{\mu + \frac{\sum_{i=1}^{q} a_i}{q}}_{\mu^*} + \underbrace{a_i - \frac{\sum_{i=1}^{q} a_i}{q}}_{a_i^*} + e_{ij}$$

$$= \mu^* + a_i^* + e_{ij}$$

$$\underline{y} = 1\underline{\mu}^* + Z\underline{a}^* + \underline{e}$$

Grunnen til denne omskriving er at elementene av vektoren $[\mu^* a_1^* a_2^* \; . \; . \; a_{q-1}^*]$ vil väre estimerbare, mens elementene av vektoren $\underline{b}' = [\mu a_1 a_2 \; . \; . \; . \; a_{q-1}]$ ikke er estimerbare (Graybill, 1961 s. 229 ; se forövrig appendiks 1). <u>Det er viktig å väre oppmerksom på at "constraint" skal väre ikke estimérbar funksjon.</u>

La 6 observasjoner for den gitte modell väre fordelt ved $i = 1, 2, 3$ og $j = 1, 2$, da vil en ha at $(\Sigma a_i = 0)$.

$$\left.\begin{aligned}
y_{11} &= \mu^* + a_1^* + e_{11} \\
y_{12} &= \mu^* + a_1^* + e_{12} \\
y_{21} &= \mu^* + a_2^* + e_{21} \\
y_{22} &= \mu^* + a_2^* + e_{22} \\
y_{31} &= \mu^* - a_1^* - a_2^* + e_{31} \\
y_{32} &= \mu^* - a_1^* - a_2^* + e_{32}
\end{aligned}\right\} \text{siden } a_3^* = -a_1^* - a_2^*$$

Dette gir:

$$\underset{\sim}{y} = \underset{\sim}{1}\mu^* + \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ -1 & -1 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} a_1^* \\ a_2^* \end{bmatrix} + \underset{=}{e}$$

som generelt kan skrives:

$$\underset{\sim}{y} = \underset{\sim}{1}\mu^* + T\underset{=}{a}^* + \underset{=}{e}$$

hvor T er en $N \times (q-1)$ matrise

$\underset{=}{a}^*$ er en $(q-1) \times 1$ vektor av de $(q-1)$ förste element av a_i^*.

Normal-ligningene for denne modellen vil väre:

$$\begin{bmatrix} 1'1 & 1'T \\ T'1 & T'T \end{bmatrix} \begin{bmatrix} \hat{\mu}^* \\ \hat{\underset{=}{a}}^* \end{bmatrix} = \begin{bmatrix} 1'y \\ T'y \end{bmatrix}$$

og estimatoren er gitt ved:

$$\begin{bmatrix} \hat{\mu}^* \\ \hat{\underset{=}{a}}^* \end{bmatrix} \begin{matrix} = \\ = \end{matrix} \begin{bmatrix} 1'1 & 1'T \\ T'1 & T'T \end{bmatrix}^{-1} \begin{bmatrix} 1'y \\ T'y \end{bmatrix}$$

$$= \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} 1'y \\ T'y \end{bmatrix} = S \begin{bmatrix} 1'y \\ T'y \end{bmatrix}$$

siden $r(S^{-1}) = q =$ dimensjonen av S^{-1}.

Vi vil her få estimater av de $(q-1)$ förste element av a^*,

dermed vil $\hat{a}_q^* = -\sum_{i=1}^{q-1} \hat{a}_i^*$

En kan derfor finne at:

$$\mathrm{Var}\ (\hat{a}_q^*) = \mathrm{Var}\ (-\sum_{i=1}^{q-1} \hat{a}_i^*)$$

$$= \mathrm{Var}\ (-\hat{a}_1^* - \hat{a}_2^* - \ldots - \hat{a}_{q-1}^*)$$

$$= \mathrm{Var}\ (\hat{a}_1^*) + \mathrm{Var}\ (\hat{a}_2^*) + \ldots + \mathrm{Var}\ (\hat{a}_{q-1}^*)$$

$$+ 2 \sum_{i<j}\sum \mathrm{Kovar}\ (\hat{a}_i^*, \hat{a}_j^*)$$

$$= 1'S_{22}1\ \sigma_e^2$$

Ved reparameterisering blir summen av henholdsvis både parametrene og estimatene lik 0. Derimot ved

$\sum_{i=1}^{q} \hat{a}_i = 0$ blir summen av estimatene lik 0.

E. Innföring av "constraints" på selve estimatorene

Dette er behandlet i kapittel V.

Appendiks 1

En vil nå diskutere den egentlige teorien som ligger bak bruken av $\sum_{i=1}^{q} a_i = 0$. Vi har at X'X er en positiv "semidefinite" matrise (Graybill, 1961 s. 3) med $r(X'X) = r = q < k$, derfor eksisterer det en "nonsingulär" matrise W med dimensjonen $k \times k$ slik at:

$$W'X'XW = \begin{bmatrix} B & 0 \\ 0 & 0 \end{bmatrix}$$

hvor B er en $r \times r$ matrise med $r(B) = r = q$ (Graybill, 1961 s. 235). Hvis vi deler W i $W = [W_1 : W_2]$, hvor W_1 er $k \times r$, får vi fölgende:

$$\begin{bmatrix} W_1' \\ W_2' \end{bmatrix} X'X\,[W_1 : W_2] = \begin{bmatrix} W_1'X'XW_1 & W_1'X'XW_2 \\ W_2'X'XW_1 & W_2'X'XW_2 \end{bmatrix} = \begin{bmatrix} B & 0 \\ 0 & 0 \end{bmatrix}$$

Dette medförer at $r(W_1'X'XW_1) = r(W_1'X') = r$, (Graybill, 1961 s. 3, teorem 1.20). Dessuten vil $W_2'X' = 0$, siden $W_2'X'XW_2 = 0$. En null-matrise som er lik og av typen $A'A(A' = W_2'X')$ kan bare väre riktig hvis A' eller A er null matriser (Graybill, 1961 s. 2).
Vi kan også skrive modellen:

$$\underline{y} = X\underline{b} + \underline{e} \text{ som } \underline{y} = XW(W)^{-1}\underline{b} + \underline{e}$$

uten å forandre den originale modellen. Hvis vi definerer:

$$W^{-1} = U = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} \quad \text{får vi:}$$

$$\begin{aligned} \underline{y} &= XWW^{-1}\underline{b} + \underline{e} \\ &= X\,[W_1 : W_2] \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} \underline{b} + \underline{e} \\ &= XW_1U_1\underline{b} + XW_2U_2\underline{b} + \underline{e} \end{aligned}$$

som reduseres til:

$$\underline{y} = XW_1U_1\underline{b} + \underline{e} \text{ , siden } XW_2 = 0.$$

Ved å definere $M = XW_1 = [1 : T]$ og $\underline{\lambda} = U_1\underline{b}$ hvor $r(M) = r = q$, kan en derfor löse normal-ligningene for:

$$M'M\underline{\lambda} = M'\underline{y}$$

Fra den definisjon av W og W^{-1} som her er gitt, har vi at:

$$WW^{-1} = I = W_1U_1 + W_2U_2$$

$$\text{mens } W^{-1}W = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} [W_1 : W_2] = \begin{bmatrix} U_1W_1 & U_1W_2 \\ U_2W_1 & U_2W_2 \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$$

så $U_1W_1 = I$, mens $M = W_1U_1 \neq I$. I denne utledning er U_1 generell så enhver funksjon $\alpha = U_1\underline{b}$ som er estimérbar kan benyttes og en av disse var vist ved det förste eksemplet (s. III 15). Den matrise U_1 som der ble benyttet er:

$$U_1 = \frac{1}{3} \begin{bmatrix} 3 & 1 & 1 & 1 \\ 0 & 2 & -1 & -1 \\ 0 & -1 & 2 & -1 \end{bmatrix}$$

Ved å multiplisere $U_1\underline{b}$ vil en finne at dette er identisk til vektoren $[\mu^* \; a_1^* \; a_2^* . \; . \; . \; a_{q-1}^*]'$ som ble definert på s. III 15.

En skal ta noen eksempler:

Först:

$$U_1\underline{b} = \frac{1}{3} \begin{bmatrix} 3 & 1 & 1 & 1 \\ 0 & 2 & -1 & -1 \\ 0 & -1 & 2 & -1 \end{bmatrix} \begin{bmatrix} \mu \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 3\mu + a_1 + a_2 + a_3 \\ 2a_1 - a_2 - a_3 \\ -a_1 + 2a_2 - a_3 \end{bmatrix}$$

• Er μ^* estimérbar? $\mu^* = \mu + \frac{1}{3} \underbrace{[a_1 + a_2 + a_3]}_{\Sigma a_i = 0} = \mu$

og derfor er μ^* estimérbar.

•• Er a_1^* estimérbar? $\frac{1}{3}$ $[2a_1 - a_2 + a_3] = \frac{1}{3} [3\, a_1 - (a_1 + a_2 + a_3)] =$

$a_1 - \underbrace{\frac{\Sigma a_i}{3}}_{0}$, og derfor er a_1^* estimérbar, osv.

Generelt vil en ha at:

$$U_1 = \frac{1}{q}\begin{bmatrix} q & 1 & 1 & \cdot\cdot & 1 & 1 \\ 0 & q-1 & -1 & \cdot\cdot & -1 & -1 \\ \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & \cdot & & & \cdot \\ 0 & -1 & -1 & & q-1 & -1 \end{bmatrix}$$

En har derfor at forventningen av estimatoren er:

$$\begin{aligned} E(\hat{\underline{\lambda}}) &= E(M'M)^{-1}M'\underline{y} \\ &= (M'M)^{-1}M'E(\underline{y}) \\ &= (M'M)^{-1}M'M\underline{\lambda} \\ &= \begin{bmatrix} \mu^* \\ \underline{a}^* \end{bmatrix} = \begin{bmatrix} \mu + \bar{a} \\ \underline{a} - \underline{1}\bar{a} \end{bmatrix} \end{aligned}$$

hvor $\bar{a} = \sum_{i=1}^{q} a_i/q$ og $\underline{1}$ er en $(q-1) \times 1$ vektor av et tall.

Derfor er estimatene forventningsrett for den forhåndsbestemte lineäre kombinasjon av parametrene.

For toveisklassifisering, hvor der er p og q klasser i hver, vil U_1 bli:

$$U_1 = \begin{bmatrix} 1 & \frac{1}{p} & \cdots & \frac{1}{p} & \frac{1}{p} & \frac{1}{q} & \cdots & \frac{1}{q} \\ 0 & \frac{p-1}{p} & \cdots & -\frac{1}{p} & -\frac{1}{p} & 0 & \cdots & 0 \\ 0 & -\frac{1}{p} & \frac{p-1}{p} \cdot & -\frac{1}{p} & -\frac{1}{p} & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & \cdot & & \cdot \\ 0 & -\frac{1}{p} & \cdots & \frac{p-1}{p} & -\frac{1}{p} & 0 & \cdots & 0 \\ 0 & \cdot\cdot & & \cdot & 0 & \frac{q-1}{q} \; -\frac{1}{q} & \cdot\cdot & -\frac{1}{q} \\ \cdot & & & & \cdot & -\frac{1}{q} \; \frac{q-1}{q} & & -\frac{1}{q} \\ \cdot & & & & \cdot & \cdot & \cdot & \cdot \\ \cdot & & & & & & \cdot & \cdot \\ \cdot & & & & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdots & & & 0 & -\frac{1}{q} & \frac{q-1}{q} & -\frac{1}{q} \end{bmatrix}$$

IV. DEFINISJON AV "BEST PREDICTOR" (BP), "BEST LINEAR PREDICTOR" (BLP) OG "BEST LINEAR UNBIASED PREDICTOR" (BLUP)

1. Prediktorer som minimaliserer "mean square error"

La oss tenke oss at vi har en registrering y_1, som vi skal benytte for å estimere ("predict") y_2. Vi forutsetter at y_1 og y_2 har samme fordeling med varians- kovarians matrisen lik $V\sigma^2$, videre at $E(y_1) = \mu_1$ og $E(y_2) = \mu_2$. Fölgelig er y_1 prediktor og y_2 prediktand. Vi vil med andre ord ha en prediktor med samme forventning som prediktanden. Hvilke kriterier skal vi legge til grunn ved valg av prediktor? I fölge Schaeffer (1975) kan en summere opp dette på fölgende måte:

a. Den beste prediktor vil väre y_2 selv, men denne er jo ukjent.

b. En annen mulighet ville väre å la $\hat{y}_2 = \mu_2$, men da må en kjenne μ_2 uten feil.

c. Hvis μ_2 er ukjent, kan vi forutsette at $\mu_1 = \mu_2$ og så la $\hat{y}_2 = y_1$

d. Et logisk krav til prediktor er at $E|\hat{y}_2 - y_2|^k$ er minimum. Vanligvis settes $k = 2$ hvilket innebärer at prediktor minimaliserer "mean square error of prediction" (MSE).

Det kan bevises at for en hvilken som helst fordeling av y_1 og y_2, er den prediktor som minimaliserer MSE lik: $\hat{y}_2 = E(y_2|y_1)$ (leses: forventningen av y_2 gitt y_1).

Om y_1 og y_2 er normalfordelte og en kjenner μ_1, μ_2, σ_{12} og σ_1^2 blir:

$$\hat{y}_2 = E(y_2|y_1) = \mu_2 + \frac{\sigma_{12}}{\sigma_1^2}(y_1 - \mu_1)$$

Her er: σ_{12} = kovariansen mellom y_1 og y_2

σ_1^2 = variansen av y_1

Under disse forutsetningene er $\hat{y}_2$ forventningsrett, og en lineär funksjon av y_1. En sier at $\hat{y}_2$ er BP.

Kjenner vi imidlertid ikke fordelingen av y_1 og y_2, kan vi begrense oss til lineäre prediktorer. I dette tilfellet er også:

$$\hat{y}_2 = \mu_2 + \frac{\sigma_{12}}{\sigma_1^2}(y_1 - \mu_1) \quad \text{, hvor}$$

$\hat{y}_2$ er forventningsrett, og minimaliserer MSE. Her er $\hat{y}_2$ BLP. En bör merke seg her at μ_1, μ_2, σ_{12} og σ_1^2 forutsettes kjent.

En ytterligere vanskelighet oppstår om bare varianser og kovarianser er kjente. Da er vi nödt til å forutsette at $\mu_1 = \mu_2 = \mu$ [el. $\mu_1 = \mu_2 = 0$], og så estimere μ ved bruk av "Generalized Least Squares" (GLS). Videre må vi kreve at $E(\hat{y}_2) = E(y_2)$. Således, $\hat{y}_2 = \hat{\mu} + \frac{\sigma_{12}}{\sigma_1^2}(y_1 - \hat{\mu})$, hvor $\hat{\mu}$ = GLS - estimator av μ.

Her er $\hat{y}_2$ forventningsrett, en lineär funksjon av y_1, og $\hat{y}_2$ minimaliserer MSE. En sier at $\hat{y}_2$ er BLUP.

2. BLUP for elementer i lineäre modeller

La oss gå ut fra fölgende lineäre modell:

$$\underline{y} = X\underline{b} + Z\underline{u} + \underline{e} \quad \text{, hvor}$$

$\underline{y}$ er en n x 1 vektor av observasjoner

$\underline{b}$ er en p x 1 vektor av ukjente konstanter ("fixed effects")

$\underline{u}$ er en q x 1 vektor av ikke-observerbare variabler ("random effects")

$\underline{e}$ er en "random" vektor av ikke-observerbare variabler forbundet med hver registrering (n x 1 vektor)

X[n x p matrise med rank = $r \leq$ min (n, p)] og Z (n x q matrise) er to matriser, som beskriver sammenhengen mellom $\underline{y}$, $\underline{b}$ og $\underline{u}$.

En forutsetter at (E = "expectation"):

$$E\begin{bmatrix} \underline{y} \\ \underline{u} \\ \underline{e} \end{bmatrix} = \begin{bmatrix} X\underline{b} \\ 0 \\ 0 \end{bmatrix} \qquad \text{og}$$

$$\text{Var.-Kov.}\begin{bmatrix} \underline{y} \\ \underline{u} \\ \underline{e} \end{bmatrix} = \begin{bmatrix} V & ZG & R \\ GZ' & G & 0 \\ R & 0 & R \end{bmatrix} \sigma^2$$

Således er f.eks.:

$$E(\underline{y}\underline{y}') - (E\underline{y})^2 =$$

$$E[(X\underline{b} + Z\underline{u} + \underline{e})(X\underline{b} + Z\underline{u} + \underline{e})'] - (X\underline{b})(X\underline{b})' =$$

$$E[X\underline{b}\underline{b}'X' + Z\underline{u}\underline{b}'X' + \underline{e}\underline{b}'X' + X\underline{b}\underline{u}'Z' + Z\underline{u}\underline{u}'Z' +$$
$$\underline{e}\underline{u}'Z' + X\underline{b}\underline{e}' + Z\underline{u}\underline{e}' + \underline{e}\underline{e}'] - [X\underline{b}\underline{b}'X'] =$$

$$E(Z\underline{u}\underline{u}'Z') + E(\underline{e}\underline{e}') = (ZGZ' + R)\sigma^2 = V\sigma^2$$

La oss ta et eksempel. Vi forutsetter 3 hanndyr med henholdsvis 3 dötre, 2 dötre og 1 datter. Fareffekten betraktes som random. Modellen blir da:

$$y_{ij} = \mu + s_i + e_{ij}$$

$s_i = \text{NID}(0, \sigma_s^2)$
$e_{ij} = \text{NID}(0, \sigma^2)$
$E(s, e) = 0$
$i = 1, 2, 3$

Bruker vi "mixed" modell får vi:

$$\underline{y} = X\underline{b} + Z\underline{u} + \underline{e}\ , \text{ hvor}$$

$\underline{y}' = (y_{11}\ y_{12}\ y_{13}\ y_{21}\ y_{22}\ y_{31})$
$\underline{b} = \mu$

$\underline{u}' = (s_1\ s_2\ s_3)$

$\underline{e}' = (e_{11}\ e_{12}\ e_{13}\ e_{21}\ e_{22}\ e_{31})$

$X' = (1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1)$

$$Z' = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$E(\underline{u}\underline{u}') = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \sigma_s^2 = I_3\sigma_s^2 = G\sigma^2$$

Således: $G = I_3 \dfrac{\sigma_s^2}{\sigma^2}$

$$R = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} = I_6$$

Modellen kunne også ha värt skrevet som "fixed" modell, dvs.:

$\underline{y} = X\underline{b} + \underline{\Sigma}$, hvor $\underline{\Sigma} = Z\underline{u} + \underline{e}$,

Var.-Kov. $(\underline{\Sigma}) = V\sigma^2$. Som kjent kan en da finne "maximum likelihood" (m.l.) estimatoren av $\underline{b}$ (eller "generalized least-squares estimator" ved ikke spesifisert fordeling av $\underline{\Sigma}$), her betegnet som $\hat{\underline{b}}$ når $\underline{\Sigma}$ er normalfordelt, på fölgende sett:

$$X'V^{-1}X\tilde{\underline{b}} = X'V^{-1}\underline{y}$$

$$\tilde{\underline{b}} = (X'V^{-1}X)^{-1}X'V^{-1}\underline{y}$$

En forutsetter at koeffisientmatrisen er ikke-singulär, og at $\underline{b}$ er estimérbar. Videre er $\underline{k}'\tilde{\underline{b}}$ BLUE ("Best linear unbiased estimator") av $\underline{k}'\underline{b}$.

Vanskeligheten ved å bruke den sistnevnte metoden er at V kan bli stor og ikke-diagonal og dermed vanskelig å invertere.

Om vi derimot bruker modellen, $\underline{y} = X\underline{b} + Z\underline{u} + \underline{e}$, og så deriverer med hensyn til $\underline{b}$ og $\underline{u}$, får vi fölgende (Henderson, 1959):

$$f(\underline{y},\underline{u}) = g(\underline{y}|\underline{u})\,h(\underline{u}) = \text{konst.} \times \exp.\, [$$

$$- \frac{1}{2\sigma^2} (\underline{y} - X\underline{b} - Z\underline{u})'R^{-1}(\underline{y} - X\underline{b} - Z\underline{u}) - \frac{1}{2\sigma^2}\underline{u}'G\underline{u}\,]$$

$$\frac{\delta f(\underline{y},\underline{u})}{\delta \underline{b}} = X'R^{-1}X\hat{\underline{b}} + X'R^{-1}Z\hat{\underline{u}} - X'R^{-1}\underline{y} = 0$$

$$\frac{\delta f(\underline{y},\underline{u})}{\delta \underline{u}} = Z'R^{-1}X\hat{\underline{b}} + (Z'R^{-1}Z + G^{-1})\hat{\underline{u}} - Z'R^{-1}\underline{y} = 0$$

som gir: $$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\underline{b}} \\ \hat{\underline{u}} \end{bmatrix} = \begin{bmatrix} X'R^{-1}\underline{y} \\ Z'R^{-1}\underline{y} \end{bmatrix}$$

Dette er med unntagelse for G^{-1} ligningene en ville bruke for å finne m.l. estimatene av $\underline{b}$ og $\underline{u}$ når $\underline{u}$ betraktes som "fixed".

Både R og G er ofte diagonalmatriser, og $Z'R^{-1}Z + G^{-1}$ er enten diagonalmatrise eller har en stor diagonal submatrise. Dette gjör selvsagt inverteringen enklere. Et spörsmål som det er naturlig å reise nå, er om $\tilde{\underline{b}} = \hat{\underline{b}}$ og i så fall er $\underline{k}'\hat{\underline{b}}$ BLUE. La oss gå ut fra den andre ligningen ovenfor.

$$Z'R^{-1}X\hat{\underline{b}} + (Z'R^{-1}Z + G^{-1})\hat{\underline{u}} = Z'R^{-1}\underline{y}$$

$$\hat{\underline{u}} = (Z'R^{-1}Z + G^{-1})^{-1}(Z'R^{-1}\underline{y} - Z'R^{-1}X\hat{\underline{b}}).$$

Vi setter dette inn i förste ligningen:

$$X'R^{-1}X\hat{\underline{b}} + X'R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}(Z'R^{-1}\underline{y} - Z'R^{-1}X\hat{\underline{b}}) =$$

$X'R^{-1}\underline{y}$, eller

$X'WX\hat{\underline{b}} = X'W\underline{y}$, hvor

$$W = R^{-1} - R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1}.$$

Om vi kan bevise at: $W = V^{-1}$, eller $VW = I$ blir $\tilde{\underline{b}} = \hat{\underline{b}}$.

$$VW = (R + ZGZ')[R^{-1} - R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1}] =$$

$$I + ZGZ'R^{-1} - Z(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1} -$$

$$ZGZ'R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1} =$$

$$I + ZGZ'R^{-1} - Z(I + GZ'R^{-1}Z)(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1} =$$

$$I + ZGZ'R^{-1} - ZG(G^{-1} + Z'R^{-1}Z)(Z'R^{-1}Z + G)^{-1}Z'R^{-1} =$$

$$I + ZGZ'R^{-1} - ZGIZ'R^{-1} = I.$$

Således er $W = V^{-1}$, og fölgelig er $\underline{k}'\hat{\underline{b}}$ BLUE av $\underline{k}'\underline{b}$.

Videre er (Henderson, 1963):

$$\hat{\underline{u}} = [Z'R^{-1}Z + G^{-1}]^{-1}[Z'R^{-1}\underline{y} - Z'R^{-1}X\hat{\underline{b}}] =$$

$$[Z'R^{-1}Z + G^{-1}]^{-1}Z'R^{-1}(\underline{y} - X\hat{\underline{b}}) = GZ'V^{-1}(\underline{y} - X\hat{\underline{b}}), \text{ fordi}$$

$$[(Z'R^{-1}Z + G^{-1})]^{-1}Z'R^{-1} = (Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1}VV^{-1} =$$

$$(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1}(ZGZ' + R)V^{-1} =$$

$$(Z'R^{-1}Z + G^{-1})^{-1}(Z'R^{-1}ZGZ' + Z')V^{-1} =$$

$$(Z'R^{-1}Z + G^{-1})^{-1}(Z'R^{-1}Z + G^{-1})GZ'V^{-1} = GZ'V^{-1}.$$

$\hat{\underline{u}}$ er BLUP, og $\underline{k}'\hat{\underline{b}} + \underline{m}'\hat{\underline{u}}$ er BLUP av $\underline{k}'\underline{b} + \underline{m}'\underline{u}$ forutsatt at $\underline{k}'\hat{\underline{b}}$ er estimérbar.

I indekssammenheng kan $\hat{\underline{u}}$ betraktes som et estimat på avlsverdien til et dyr, og höyre siden av normal-ligningene er GZ'.

Et <u>annet</u> sett å vise at $\underline{k}'\hat{\underline{b}}$ er BLUE, $\hat{\underline{u}}$ er BLUP og $\underline{k}'\hat{\underline{b}} + \underline{m}'\hat{\underline{u}}$ er BLUP hvor ^ er lösningene ved bruk av "mixed model"-ligningene, er som fölger (Schaeffer, 1975):

La oss tenke oss en vektor av observasjoner, $\underline{y}$, eksempelvis registrering på far, mor og avkom. Denne vektoren vil vi bruke til å predikte $\underline{w}$, en ikke observérbar vektor. $\underline{w}$ kan f.eks. väre avlsverdien til fremtidig avkom. Dette betyr at om:

$$E \begin{bmatrix} \underline{y} \\ \underline{w} \end{bmatrix} = \begin{bmatrix} X\underline{b} \\ P\underline{b} \end{bmatrix}$$

, blir den funksjonen vi vil predikte lik $k'\underline{b} + m'\underline{w}$.

Vårt krav til prediksjonen er at:

$E(\underline{f}'\underline{y}) = E(\underline{k}'\underline{b} + \underline{m}'\underline{w})$ dvs. "unbiasedness" og

$E(\underline{f}'\underline{y} - \underline{k}'\underline{b} - \underline{m}'\underline{w})^2$ = "Mean square error" (MSE) er minimalisert.

For å oppfylle disse kravene må en kjenne til varians- kovariansmatrisen, dvs.

$$\text{Var.-Kov.} \begin{bmatrix} \underline{y} \\ \underline{w} \end{bmatrix} = \begin{bmatrix} V & C \\ C' & H \end{bmatrix} \sigma^2$$

Den funksjonen som skal minimaliseres blir da:

$$g = \underbrace{\underline{f}'V\underline{f} + \underline{m}'H\underline{m} - 2\underline{f}'C\underline{m}}_{E(\underline{f}'\underline{y} - \underline{k}'\underline{b} - \underline{m}'\underline{w})^2} + \underbrace{(\underline{f}'X - \underline{k}' - \underline{m}'P)\underline{\phi}}_{\text{Sikrer "unbiasedness"}}$$

hvor $\underline{\phi}$ er en vektor av "LaGrange multipliers", som brukes for å få forventningsretthet. En minimaliserer så funksjonen g med hensyn til $\underline{f}$ og $\underline{\phi}$.

$$\frac{\delta g}{\delta f} = V\underline{f} + X\tfrac{1}{2}\underline{\phi} - C\underline{m} = 0$$

$$\frac{\delta g}{\delta \phi} = X'\underline{f} + 0\tfrac{1}{2}\underline{\phi} - (\underline{k} + P\underline{m}) = 0$$

hvilket gir:

$$\begin{bmatrix} V & X \\ X' & 0 \end{bmatrix} \begin{bmatrix} \underline{f} \\ \frac{1}{2}\underline{\phi} \end{bmatrix} = \begin{bmatrix} C\underline{m} \\ \underline{k} + P\underline{m} \end{bmatrix}$$

Etter å ha beregnet f og $\frac{1}{2}\phi$, kan en finne prediktoren $\underline{f}'\underline{y}$ (se nedenfor).

Om vi går ut fra at F og θ er matriser, blir prediktanden lik:

$K'\underline{b} + M'\underline{w}$, og

$$\begin{bmatrix} V & X \\ X' & 0 \end{bmatrix} \begin{bmatrix} F \\ \theta \end{bmatrix} = \begin{bmatrix} CM \\ K + PM \end{bmatrix}$$

Löser vi disse ligningene finner vi at:

$VF + X\theta = CM$ som gir $\theta = X^{-}(CM - VF)$

Videre, $X'F = K + PM$ som gir $F = (X')^{-}(K + PM)$.

Löser en så med hensyn på θ, får vi:

$$\theta = (X'V^{-1}X)^{-} \; (X'V^{-1}CM - K - PM).$$

Et annet uttrykk for F er: $F = V^{-1}CM - V^{-1}X\theta$, og prediktoren blir da:

$$F'\underline{y} = M'C'V^{-1}\underline{y} - (M'C'V^{-1}X - K' - M'P')(X'V^{-1}X)^{-}X'V^{-1}\underline{y} =$$

$$(K' + M'P')\hat{\underline{b}} + M'C'V^{-1}(\underline{y} - X\hat{\underline{b}})$$

hvor:

hvor:

$\hat{\underline{b}} = (X'V^{-1}X)^{-} X'V^{-1}\underline{y}$, et GLS estimat av $\underline{b}$. Når $(X'V^{-1}X)$ er ikke-singulär, blir:

$\hat{\underline{b}} = (X'V^{-1}X)^{-1}X'V^{-1}\underline{y}$, altså den samme lösningen som tidligere.

Hva med $\hat{\underline{u}}$?

Om $\underline{u}$ = w, K' = 0 og M' = I, blir:

$K'\underline{b} + M'w = \underline{u}$ (prediktand) og

$F'\underline{y} = \hat{u} = P'\hat{\underline{b}} + C'V^{-1}(\underline{y} - X\hat{\underline{b}})$.

Nå er $E(\underline{u}) = \underline{0}$. Dette innebärer at $E(\underline{w}) = P\underline{b} = 0$ og dermed P = 0 (null-matrise) for alle verdier av $\underline{b}$. Derfor:

$\hat{\underline{u}} = C'V^{-1}(\underline{y} - X\hat{\underline{b}})$, altså er $\hat{\underline{u}}$ beregnet med "mixed model" ligningene BLUP.

En ytterligere komplisering oppstår om $E(\hat{\underline{u}}) \neq 0$, $E(\hat{\underline{e}}) \neq 0$ og $kov(\underline{u},\underline{e}) = S \neq \underline{0}$. En får da (Schaeffer, 1975):

$$E \begin{bmatrix} \underline{y} \\ \underline{u} \\ \underline{e} \end{bmatrix} = \begin{bmatrix} X\underline{b} + ZP\underline{b} + T\underline{b} \\ P\underline{b} \\ T\underline{b} \end{bmatrix} \text{, og}$$

$$\text{Var.-Kov.} \begin{bmatrix} \underline{y} \\ \underline{u} \\ \underline{e} \end{bmatrix} = \begin{bmatrix} V & ZG + S' & ZS + R \\ & G & S \\ \text{symmetrisk} & & R \end{bmatrix} \text{, hvor}$$

V = ZGZ' + ZS + S'Z' + R. Det kan bevises at om

$R - S'G^{-1}S = A$ og om G^{-1} og R^{-1} eksisterer, så er:

$A^{-1} = R^{-1} - R^{-1}S'(SR^{-1}S' - G)^{-1}SR^{-1}$

$V^{-1} = A^{-1} - A^{-1}(ZG + S')W^{-1}(GZ' + S)A^{-1}$

Her er $W = (GZ' + S)A^{-1}(ZG + S') + G$

Om $\Gamma = X + ZP + T$ blir "mixed model"-ligningene:

$$\begin{bmatrix} \Gamma'A^{-1}\Gamma & \Gamma'A^{-1}(ZG + S') \\ (S + GZ')A^{-1}\Gamma & (S + GZ')A^{-1}(ZG + S') + G \end{bmatrix} \begin{bmatrix} \hat{\underline{b}} \\ \hat{\underline{t}} \end{bmatrix} = \begin{bmatrix} \Gamma'A^{-1}\underline{y} \\ (GZ' + S)A^{-1}\underline{y} \end{bmatrix}$$

hvor $\hat{\underline{u}} = G\hat{\underline{t}} + P\hat{\underline{b}}$.

En kan merke seg at disse ligningene egentlig innebärer modellen:

$$\underline{y} = X\underline{b} + (ZG + S')\underline{t} + \underline{k}$$

3. Oppsummering av karakteristikker ved "mixed model"-ligningene

Vi skal så summere opp noen karakteristikker ved "mixed model"-ligningene (Henderson, 1973, 1975 a):

- $\underline{k}'\hat{\underline{b}}$ er BLUE av $\underline{k}'\underline{b}$

•• $\hat{\underline{u}}$ er BLUP. Dette innebärer bl.a. at under forutsetning av normalfordeling er:

$E(\underline{u}|\hat{\underline{u}})$ = (det betingete gjennomsnitt av $\underline{u}$ ved gitt $\hat{\underline{u}}$ =) $\hat{\underline{u}}$

Var $(\hat{\underline{u}} - \underline{u})$ = Var $(\underline{u}|\hat{\underline{u}})$

$\hat{\underline{u}}$ maksimerer sannsynligheten for riktig rangering av elementene av $\underline{u}$ når en betrakter lineäre prediktorer med gjennomsnitt lik 0. Det er av interesse å vite at BLUP i indekssammenheng maksimerer korrelasjonen mellom avlsverdien, T (prediktand), og seleksjonsindeksen, I (prediktor), og dermed minimaliserer MSE når $\sigma_{TI} = \sigma_I^2$, maksimerer sannsynligheten for korrekt rangering av T basert på I og maksimerer genetisk fremgang ved vertikal avskjäring av fordelingskurven.

••• Lösningen for $\hat{\underline{u}}$ er unik uansett om koeffisientmatrisen for "mixed model"-ligningene er av full rank eller ikke.

•••• $\underline{k}'\hat{\underline{b}} + \underline{m}'\hat{\underline{u}}$ er BLUP av $\underline{k}'\underline{b} + \underline{m}'\underline{u}$ forutsatt at $\underline{k}'\hat{\underline{b}}$ er estimérbar.

..... La $\begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$ väre en "generalized inverse" for

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix}$$

C_{22} er alltid unik mens C_{11} og C_{12} er unike bare når X er av full rank. Vi får da:

$$\text{Var. } (K'\hat{\underline{b}}) = K'C_{11}K\sigma^2 \text{ når } K'\hat{\underline{b}} \text{ er estimérbar.}$$

$$\text{Kov. } (K'\hat{\underline{b}},\hat{\underline{u}}') = 0$$

$$\text{Kov. } (K'\hat{\underline{b}},\underline{u}') = -K'C_{12}\sigma^2$$

$$\text{Kov. } (K'\hat{\underline{b}},\hat{\underline{u}}' - \underline{u}') = K'C_{12}\sigma^2$$

$$\text{Var. } (\hat{\underline{u}}) = \text{Kov. } (\hat{\underline{u}},\underline{u}') = (G - C_{22})\sigma^2$$

$$\text{Var. } (\hat{\underline{u}} - \underline{u}') = C_{22}\sigma^2$$

...... Om vi uttrykker lösningene for $K'\hat{\underline{b}}$ og $\hat{\underline{u}}$ ut fra elementene i modellen blir:

$$K'\hat{\underline{b}} = K'C_{11}X'R^{-1}\underline{y} + K'C_{12}Z'R^{-1}\underline{y} =$$

$$K'C_{11}X'R^{-1}(X\underline{b} + Z\underline{u} + \underline{e}) + K'C_{12}Z'R^{-1}(X\underline{b} + Z\underline{u} + \underline{e}) =$$

$$K'C_{11}X'R^{-1}X\underline{b} + K'C_{11}X'R^{-1}Z\underline{u} + K'C_{11}X'R^{-1}\underline{e} =$$

$$K'C_{12}Z'R^{-1}X\underline{b} + K'C_{12}Z'R^{-1}Z\underline{u} + K'C_{12}Z'R^{-1}\underline{e} =$$

$$K'(C_{11}X'R^{-1}X + C_{12}Z'R^{-1}X)\underline{b} +$$
$$K'(C_{11}X'R^{-1}Z + C_{12}Z'R^{-1}Z)\underline{u} +$$
$$K'(C_{11}X'R^{-1} + C_{12}Z'R^{-1})\underline{e} =$$

$$K'\underline{b} + K'(\underbrace{C_{11}X'R^{-1}Z + C_{12}Z'R^{-1}Z - C_{12}G^{-1}}_{0} + C_{12}G^{-1})\underline{u} +$$

$$+ K'(C_{11}X' + C_{12}Z')R^{-1}\underline{e} =$$

$$K'\underline{b} + K'C_{12}G^{-1}\underline{u} + K'(C_{11}X' + C_{12}Z')R^{-1}\underline{e}$$

På samme sett kan en bevise at:

$$\hat{\underline{u}} = (I - C_{22}G^{-1})\underline{u} + (C_{21}X' + C_{22}Z')R^{-1}\underline{e}$$

V. BLUP (BLUE)-METODEN ILLUSTRERT MED NUMERISKE EKSEMPLER

I det fölgende skal en illustrere fleksibiliteten i BLUP-metoden ved praktiske eksempler.

EKSEMPEL A. "RANDOM" EFFEKTER - "FIXED" EFFEKTER - ESTIMERBARHET - VARIANSER (VAN VLECK, 1975)

Vi forutsetter fölgende datasett hvor fedrene A, B og C hörer til gruppe 1, og fedrene D og E hörer til gruppe 2. Tallene utenfor parentes angir antall dötre mens tallene inne i parentesen angir summen av registreringene:

Gruppe	Fedre	Besetning-år-sesong 1	2	3	Totalt antall registre-ringer		Sum av registre-ringene	
1	A = 11	2 (28)	- --	1 (11)	3		39	
	B = 12	- --	1 (13)	3 (40)	4		53	
	C = 13	- --	2 (30)	- --	2	9	30	(122)
2	D = 21	2 (22)	3 (45)	- --	5		67	
	E = 22	- --	4 (40)	5 (55)	9	14	95	(162)
Totalt antall registre-ringer		4	10	9	23			
Sum av registre-ringene		50	128	106				

Modellen er:

$$y_{ijkl} = \mu + h_i + g_j + s_{jk} + e_{ijkl} ,$$

hvor

μ = konstant

h_i = effekt av besetning-år-sesong

g_j = effekt av j^{te} fargruppe

s_{jk} = effekt av k^{te} far innen j^{te} fargruppe

e_{ijkl} = en tilfeldig effekt forbundet med registreringen på l^{te} avkom etter k^{te} far innen j^{te} gruppe og i^{te} besetning-år-sesong $[0,1]$

y_{ijkl} = observasjon på l^{te} avkom etter k^{te} far innen j^{te} gruppe og i^{te} besetning-år-sesong

a. $\underline{h}$, $\underline{g}$ og $\underline{s}$ er "fixed" effekter

Modellen kan da uttrykkes som:

$$\underline{y} = X\underline{b} + \underline{e} \quad \text{og} \quad E(\underline{y}) = X\underline{b}$$

X og $\underline{b}$ er satt opp på side V A:3.

Videre:

$$X'X\underline{b} = X'\underline{y} \text{ , som for dette eksemplet blir:}$$

$$
\begin{array}{ccccccccccc}
\underline{\mu} & \underline{h_1} & \underline{h_2} & \underline{h_3} & \underline{g_1} & \underline{g_2} & \underline{s_{11}} & \underline{s_{12}} & \underline{s_{13}} & \underline{s_{21}} & \underline{s_{22}} \\
n_{...} & n_{1..} & n_{2..} & n_{3..} & n_{.1.} & n_{.2.} & n_{.11} & n_{.12} & n_{.13} & n_{.21} & n_{.22}
\end{array}
$$

$$
\begin{bmatrix}
23 & 4 & 10 & 9 & 9 & 14 & 3 & 4 & 2 & 5 & 9 \\
 & 4 & 0 & 0 & 2 & 2 & 2 & 0 & 0 & 2 & 0 \\
 & & 10 & 0 & 3 & 7 & 0 & 1 & 2 & 3 & 4 \\
 & & & 9 & 4 & 5 & 1 & 3 & 0 & 0 & 5 \\
 & & & & 9 & 0 & 3 & 4 & 2 & 0 & 0 \\
 & & & & & 14 & 0 & 0 & 0 & 5 & 9 \\
\text{symmetrisk} & & & & & & 3 & 0 & 0 & 0 & 0 \\
 & & & & & & & 4 & 0 & 0 & 0 \\
 & & & & & & & & 2 & 0 & 0 \\
 & & & & & & & & & 5 & 0 \\
 & & & & & & & & & & 9
\end{bmatrix}
\begin{bmatrix}
\hat{\mu} \\ \hat{h}_1 \\ \hat{h}_2 \\ \hat{h}_3 \\ \hat{g}_1 \\ \hat{g}_2 \\ \hat{s}_{11} \\ \hat{s}_{12} \\ \hat{s}_{13} \\ \hat{s}_{21} \\ \hat{s}_{22}
\end{bmatrix}
=
\begin{bmatrix}
\overset{y_{....}}{284} \\ \overset{y_{1...}}{50} \\ \overset{y_{2...}}{128} \\ \overset{y_{3...}}{106} \\ \overset{y_{.1..}}{122} \\ \overset{y_{.2..}}{162} \\ \overset{y_{.11.}}{39} \\ \overset{y_{.12.}}{53} \\ \overset{y_{.13.}}{30} \\ \overset{y_{.21.}}{67} \\ \overset{y_{.22.}}{95}
\end{bmatrix}
$$

Denne matrisen er ikke av full rank, og en må derfor innföre fire* "constraints" (dvs. ikke-estimérbare funksjoner).

$$
\begin{array}{c}
\qquad\qquad\qquad\qquad X \\
\begin{array}{ccccccccccc}
\mu & h_1 & h_2 & h_3 & g_1 & g_2 & s_{11} & s_{12} & s_{13} & s_{21} & s_{22}
\end{array} \\
\begin{bmatrix}
1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
\end{array}
\qquad
\begin{array}{c}
\underline{b} \\
\mu \\
\begin{bmatrix}
h_1 \\ h_2 \\ h_3 \\ g_1 \\ g_2 \\ s_{11} \\ s_{12} \\ s_{13} \\ s_{21} \\ s_{22}
\end{bmatrix}
\end{array}
$$

*Koeffisientene for henholdsvis $\underline{h}$ og $\underline{g}$ summerer opp til koeffisientene for μ. Videre koeffisientene for s_{11}, s_{12} og s_{13} summerer opp til koeffisientene for g_1, og koeffisientene for s_{21} og s_{22} summerer opp til koeffisientene for g_2

La oss velge fölgende "constraints":

$$\sum_i \hat{h}_i = \sum_j \hat{g}_j = \sum_k \hat{s}_{jk} = 0 \quad (j = 1,\ 2)$$

Rent teknisk sett kan en innföre disse 4 "constraints" på flere måter. Harvey (1960) bruker subtraksjonsmetoden mens en her velger å bruke "LaGrange multipliers" dvs. en utöker $X'X\underline{b} = X'\underline{y}$ på fölgende sett:

$$\begin{array}{ccccccccccc} \underline{\mu} & \underline{h_1} & \underline{h_2} & \underline{h_3} & \underline{g_1} & \underline{g_2} & \underline{s_{11}} & \underline{s_{12}} & \underline{s_{13}} & \underline{s_{21}} & \underline{s_{22}} \\ n_{\cdots} & n_{1\cdot\cdot} & n_{2\cdot\cdot} & n_{3\cdot\cdot} & n_{\cdot 1\cdot} & n_{\cdot 2\cdot} & n_{\cdot 11} & n_{\cdot 12} & n_{\cdot 13} & n_{\cdot 21} & n_{\cdot 22} \end{array}$$

$$\begin{bmatrix}
23 & 4 & 10 & 9 & 9 & 14 & 3 & 4 & 2 & 5 & 9 & 0 & 0 & 0 & 0 \\
 & 4 & 0 & 0 & 2 & 2 & 2 & 0 & 0 & 2 & 0 & 1 & 0 & 0 & 0 \\
 & & 10 & 0 & 3 & 7 & 0 & 1 & 2 & 3 & 4 & 1 & 0 & 0 & 0 \\
 & & & 9 & 4 & 5 & 1 & 3 & 0 & 0 & 5 & 1 & 0 & 0 & 0 \\
 & & & & 9 & 0 & 3 & 4 & 2 & 0 & 0 & 0 & 1 & 0 & 0 \\
 & & & & & 14 & 0 & 0 & 0 & 5 & 9 & 0 & 1 & 0 & 0 \\
\text{symmetrisk} & & & & & & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
 & & & & & & & 4 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
 & & & & & & & & 2 & 0 & 0 & 0 & 0 & 1 & 0 \\
 & & & & & & & & & 5 & 0 & 0 & 0 & 0 & 1 \\
 & & & & & & & & & & 9 & 0 & 0 & 0 & 1 \\
0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix} \hat{\mu} \\ \hat{h}_1 \\ \hat{h}_2 \\ \hat{h}_3 \\ \hat{g}_1 \\ \hat{g}_2 \\ \hat{s}_{11} \\ \hat{s}_{12} \\ \hat{s}_{13} \\ \hat{s}_{21} \\ \hat{s}_{22} \\ \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{bmatrix}
=
\begin{bmatrix} \overset{y_{\cdots}}{284} \\ \overset{y_{1\cdot\cdot}}{50} \\ \overset{y_{2\cdot\cdot}}{128} \\ \overset{y_{3\cdot\cdot}}{106} \\ \overset{y_{\cdot 1\cdot}}{122} \\ \overset{y_{\cdot 2\cdot}}{162} \\ \overset{y_{\cdot 11\cdot}}{39} \\ \overset{y_{\cdot 12\cdot}}{53} \\ \overset{y_{\cdot 13\cdot}}{30} \\ \overset{y_{\cdot 21\cdot}}{67} \\ \overset{y_{\cdot 22\cdot}}{95} \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

λ_t $(t = 1..4)$ = "LaGrange multipliers"

Fölgelig:

$$\underline{\hat{b}} = (X'X)^{-}\,X'\underline{y} \quad \text{og} \quad E(\underline{\hat{b}}) = (X'X)^{-}X'E(\underline{y}) = \underbrace{(X'X)^{-}}_{\substack{\text{"generalized inverse"}\\ \text{av de opprinnelige}\\ \text{"least-squares"-}\\ \text{ligningene}}}\underbrace{X'X}_{\substack{\text{de opprinnelige}\\ \text{"least-squares"-}\\ \text{ligningene}}}\underline{b}$$

En må altså kjenne $(X'X)^{-}$ for å finne hva som er estimérbart ved bruk av denne metoden.

For vårt eksempel:

$$\underline{\hat{b}} = \begin{bmatrix} \hat{\mu} \\ \hat{h}_1 \\ \hat{h}_2 \\ \hat{h}_3 \\ \hat{g}_1 \\ \hat{g}_2 \\ \hat{s}_{11} \\ \hat{s}_{12} \\ \hat{s}_{13} \\ \hat{s}_{21} \\ \hat{s}_{22} \end{bmatrix} = \begin{bmatrix} 12{,}672 \\ -0{,}985 \\ 0{,}570 \\ 0{,}415 \\ 0{,}910 \\ -0{,}910 \\ -0{,}063 \\ -0{,}785 \\ 0{,}898 \\ 1{,}690 \\ -1{,}690 \end{bmatrix} \qquad E[\hat{b}] = \begin{array}{c} \begin{array}{ccccccccccc} \mu & h_1 & h_2 & h_3 & g_1 & g_2 & s_{11} & s_{12} & s_{13} & s_{21} & s_{22} \end{array} \\ \begin{bmatrix} 1 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{2} & \frac{1}{2} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{4} & \frac{1}{4} \\ & \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} & & & & & & & \\ & -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} & & & & & & & \\ & -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} & & & & & & & \\ & & & & \frac{1}{2} & -\frac{1}{2} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & -\frac{1}{4} & -\frac{1}{4} \\ & & & & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{6} & -\frac{1}{6} & -\frac{1}{6} & \frac{1}{4} & \frac{1}{4} \\ \text{ikke symmetrisk} & & & & & & \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} & & \\ & & & & & & -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} & & \\ & & & & & & -\frac{1}{2} & -\frac{1}{3} & \frac{2}{3} & & \\ & & & & & & & & & \frac{1}{2} & -\frac{1}{2} \\ & & & & & & & & & -\frac{1}{2} & \frac{1}{2} \end{bmatrix} \end{array} \begin{bmatrix} \mu \\ h_1 \\ h_2 \\ h_3 \\ g_1 \\ g_2 \\ s_{11} \\ s_{12} \\ s_{13} \\ s_{21} \\ s_{22} \end{bmatrix}$$

En bör merke seg at en skal ignorére rekker og kolonner som i "generalized inverse" korresponderer til λ_t, når en tar "expectations" og estimérer varianser på effektene.

For dette eksemplet blir f.eks.:

$$E(\hat{h}_1 - \hat{h}_2) = \frac{2}{3}\,h_1 - \frac{1}{3}\,h_2 - \frac{1}{3}\,h_3 - [-\frac{1}{3}\,h_1 + \frac{2}{3}\,h_2 - \frac{1}{3}h_3] = h_1 - h_2,$$

og dermed er differansen $\hat{h}_1 - \hat{h}_2$, estimérbar.

$$E(\hat{h}_1) = \frac{2}{3}\,h_1 - \frac{1}{3}\,(h_2 + h_3) = h_1 - \frac{1}{3}\,(h_1 + h_2 + h_3)$$, hvilket en

skulle forvente.

Nedenfor har en vist den inverse koeffisientmatrisen:

$$\left[\begin{pmatrix} X'X & X'Z \\ Z'X & Z'Z \end{pmatrix}^{-} = \begin{pmatrix} C_{11} & C_{12} \\ C'_{12} & C_{22} \end{pmatrix}\right]$$

	$\hat{\mu}$	$\hat{h}_1$	$\hat{h}_2$	$\hat{h}_3$	$\hat{g}_1$	$\hat{g}_2$	$\hat{s}_{11}$	$\hat{s}_{12}$	$\hat{s}_{13}$	$\hat{s}_{21}$	$\hat{s}_{22}$
$\hat{\mu}$	0,058	0,042	-0,024	-0,013	0,009	-0,009	-0,036	-0,009	0,045	0,002	-0,002
$\hat{h}_1$		0,241	-0,106	-0,134	-0,003	0,003	-0,155	0,088	0,067	-0,077	0,077
$\hat{h}_2$			0,120	-0,013	0,008	-0,008	0,097	0,001	-0,098	0,008	-0,008
$\hat{h}_3$				0,148	-0,005	0,005	0,058	-0,090	0,031	0,069	-0,069
$\hat{g}_1$					0,050	-0,050	-0,000	-0,016	0,016	-0,013	0,013
$\hat{g}_2$						0,050	0,000	0,016	-0,016	0,013	-0,013
$\hat{s}_{11}$	symmetrisk						0,342	-0,115	-0,227	0,040	-0,040
$\hat{s}_{12}$								0,258	-0,143	-0,043	0,043
$\hat{s}_{13}$									0,371	0,003	-0,003
$\hat{s}_{21}$										0,112	-0,112
$\hat{s}_{22}$											0,112

Hva er variansen på differensen $\hat{h}_1 - \hat{h}_2$?

$$\mathrm{Var}\,(\hat{h}_1 - \hat{h}_2) = \mathrm{Var}\,(\hat{h}_1) + \mathrm{Var}\,(\hat{h}_2) - 2\,\mathrm{Kov}\,(\hat{h}_1, \hat{h}_2) =$$

$$[0{,}241 + 0{,}120 - (2)\,(-0{,}106)]\,\sigma^2 = \underline{0{,}573\sigma^2}$$

Hva er Var $[(\hat{g}_1 + \hat{s}_{11}) - (\hat{g}_2 + \hat{s}_{22})]$?

$$\mathrm{Var}\,[(\hat{g}_1 + \hat{s}_{11}) - (\hat{g}_2 + \hat{s}_{22})] = \mathrm{Var}\,(\hat{g}_1) + \mathrm{Var}\,(\hat{g}_2) - 2\,\mathrm{Kov}\,(\hat{g}_1, \hat{g}_2) + \mathrm{Var}\,(\hat{s}_{11}) + \mathrm{Var}\,(\hat{s}_{22}) - 2\,\mathrm{Kov}\,(\hat{s}_{11}, \hat{s}_{22}) - 2\,\mathrm{Kov}\,(\hat{g}_1, \hat{s}_{22}) - 2\,\mathrm{Kov}\,(\hat{g}_2, \hat{s}_{11}) + 2\,\mathrm{Kov}\,(\hat{g}_1, \hat{s}_{11}) + 2\,\mathrm{Kov}\,(\hat{g}_2, \hat{s}_{22}) =$$

$$[0{,}050 + 0{,}050 - (2)(-0{,}050) + (0{,}342) + (0{,}112) - (2)(-0{,}040) - (2)(0{,}013) - (2)(0{,}000) + (2)(-0{,}000) + (2)(-0{,}013)]\sigma^2 = \underline{0{,}682\sigma^2}$$

Om en derimot velger fölgende fire "constraints";
$\hat{\mu} = 0$; $\hat{h}_1 = 0$; $\hat{g}_1 = \hat{g}_2 = 0$, blir:

$$\hat{\underline{b}} = \begin{bmatrix} \hat{\mu} \\ \hat{h}_1 \\ \hat{h}_2 \\ \hat{h}_3 \\ \hat{g}_1 \\ \hat{g}_2 \\ \hat{s}_{11} \\ \hat{s}_{12} \\ \hat{s}_{13} \\ \hat{s}_{21} \\ \hat{s}_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1{,}555 \\ 1{,}400 \\ 0 \\ 0 \\ 12{,}533 \\ 11{,}811 \\ 13{,}445 \\ 12{,}467 \\ 9{,}086 \end{bmatrix} \qquad E(\hat{\underline{b}}) = \begin{array}{c} \begin{array}{ccccccccccc} \mu & h_1 & h_2 & h_3 & g_1 & g_2 & s_{11} & s_{12} & s_{13} & s_{21} & s_{22} \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & & & & & \\ 0 & -1 & 1 & 0 & 0 & 0 & & & & & \\ 0 & -1 & 0 & 1 & 0 & 0 & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & & & & & \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & & & & \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & & & \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & & \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array} \begin{bmatrix} \mu \\ h_1 \\ h_2 \\ h_3 \\ g_1 \\ g_2 \\ s_{11} \\ s_{12} \\ s_{13} \\ s_{21} \\ s_{22} \end{bmatrix}$$

Den inverse koeffisientmatrisen er:

	$\hat{\mu}$	$\hat{h}_1$	$\hat{h}_2$	$\hat{h}_3$	$\hat{g}_1$	$\hat{g}_2$	$\hat{s}_{11}$	$\hat{s}_{12}$	$\hat{s}_{13}$	$\hat{s}_{21}$	$\hat{s}_{22}$
$\hat{\mu}$	0	0	0	0	0	0	0	0	0	0	0
$\hat{h}_1$	0	0	0	0	0	0	0	0	0	0	0
$\hat{h}_2$	0	0	0,573	0,468	0	0	-0,156	-0,494	-0,573	-0,344	-0,515
$\hat{h}_3$	0	0	0,468	0,657	0	0	-0,219	-0,610	-0,468	-0,281	-0,573
$\hat{g}_1$	0	0	0	0	0	0	0	0	0	0	0
$\hat{g}_2$	0	0	0	0	0	0	0	0	0	0	0
$\hat{s}_{11}$	0	0	-0,156	-0,219	0	0	0,406	0,203	0,156	0,094	0,191
$\hat{s}_{12}$	0	0	-0,494	-0,610	0	0	0,203	0,831	0,494	0,297	0,559
$\hat{s}_{13}$	0	0	-0,573	-0,468	0	0	0,156	0,494	1,073	0,344	0,515
$\hat{s}_{21}$	0	0	-0,344	-0,281	0	0	0,094	0,297	0,343	0,406	0,309
$\hat{s}_{22}$	0	0	-0,515	-0,573	0	0	0,191	0,559	0,515	0,309	0,658

I dette tilfellet er:

$$E\ (\hat{h}_1 - \hat{h}_2) = 0 - [-h_1 + h_2] = h_1 - h_2,$$

og

$$\text{Var}\ (\hat{h}_1 - \hat{h}_2) = [0 + 0{,}573 - 2\ (0)]\sigma^2 = \underline{0{,}573\sigma^2}$$

<u>En ser her at resultatet blir det samme uansett hvilket sett av "constraints" en innförer på estimatene.</u>

<u>b. $\underset{\sim}{h}$ og g er "fixed" effekter, men $\underset{\sim}{s}$ er "random" effekt</u>

Vi har da modellen:

$$\underset{\sim}{y} = X\underset{\sim}{b} + Z\underset{\sim}{u} + \underset{\sim}{e}$$

hvor

$$\underset{\sim}{b} \quad \begin{bmatrix} \mu \\ h_1 \\ h_2 \\ h_3 \\ g_1 \\ g_2 \end{bmatrix} \quad , \quad \underset{\sim}{u} = \begin{bmatrix} s_{11} \\ s_{12} \\ s_{13} \\ s_{21} \\ s_{22} \end{bmatrix}$$

BLUP-ligningene er:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z+G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\underset{\sim}{b}} \\ \hat{\underset{\sim}{u}} \end{bmatrix} = \begin{bmatrix} X'\underset{\sim}{y} \\ Z'\underset{\sim}{y} \end{bmatrix}$$

$R = I\sigma_e^2$; $G = I\sigma_s^2$. Om vi setter $R = I$ ("scaled"), blir:

$G = I \dfrac{\sigma_s^2}{\sigma_e^2}$. Videre $h^2 = \dfrac{4\ \sigma_s^2}{\sigma_s^2 + \sigma_c^2}$, settes så $h^2 = 0{,}25$ blir:

$\dfrac{\sigma_e^2}{\sigma_s^2} = 15.$ $E\ (\underset{\sim}{u}) = 0.$

$$
\left[\begin{array}{cccccc|ccccc}
23 & 4 & 10 & 9 & 9 & 14 & 3 & 4 & 2 & 5 & 9 \\
 & 4 & 0 & 0 & 2 & 2 & 2 & 0 & 0 & 2 & 0 \\
 & & 10 & 0 & 3 & 7 & 0 & 1 & 2 & 3 & 4 \\
 & & & 9 & 4 & 5 & 1 & 3 & 0 & 0 & 5 \\
 & & & & 9 & 0 & 3 & 4 & 2 & 0 & 0 \\
 & & & & & 14 & 0 & 0 & 0 & 5 & 9 \\
\hline
\multicolumn{6}{l|}{\text{symmetrisk}} & 3+15 & 0 & 0 & 0 & 0 \\
 & & & & & & & 4+15 & 0 & 0 & 0 \\
 & & & & & & & & 2+15 & 0 & 0 \\
 & & & & & & & & & 5+15 & 0 \\
 & & & & & & & & & & 9+15
\end{array}\right]
\left[\begin{array}{c}
\hat{\mu} \\ \hat{h}_1 \\ \hat{h}_2 \\ \hat{h}_3 \\ \hat{g}_1 \\ \hat{g}_2 \\ \hline \hat{s}_{11} \\ \hat{s}_{12} \\ \hat{s}_{13} \\ \hat{s}_{21} \\ \hat{s}_{22}
\end{array}\right]
=
\left[\begin{array}{c}
y_{....} \; 284 \\ y_{1...} \; 50 \\ y_{2...} \; 128 \\ y_{3...} \; 106 \\ y_{.1..} \; 122 \\ y_{.2..} \; 162 \\ \hline y_{.11.} \; 39 \\ y_{.12.} \; 53 \\ y_{.13.} \; 30 \\ y_{.21.} \; 67 \\ y_{.22.} \; 95
\end{array}\right]
$$

Her må vi innföre to (koeffisientene for henholdsvis $\underline{h}$ og g summerer opp til koeffisientene for μ) ikke-estimérbare "constraints", og lösningene blir:

$$
\begin{bmatrix} \hat{\underline{b}} \\ \hat{\underline{u}} \end{bmatrix}
=
\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z+G^{-1} \end{bmatrix}^{-}
\begin{bmatrix} X'\underline{y} \\ Z'\underline{y} \end{bmatrix}
=
\begin{bmatrix} C_{11} & C_{12} \\ C_{12}' & C_{22} \end{bmatrix}
\begin{bmatrix} X'\underline{y} \\ Z'\underline{y} \end{bmatrix}
$$

videre:

$$
E \begin{bmatrix} \hat{\underline{b}} \\ \hat{\underline{u}} \end{bmatrix}
=
\begin{bmatrix} C_{11} & C_{12} \\ C_{12}' & C_{22} \end{bmatrix}
\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z \end{bmatrix}
\begin{bmatrix} \underline{b} \\ \underline{u} \end{bmatrix}
$$

NB! G^{-1} er ignorert her

Om vi velger fölgende "constraints" ; $\hat{\mu} = 0$; $\hat{g}_1 = 0$, blir:

$$\begin{bmatrix}\hat{\underline{b}}\\ \hat{\underline{u}}\end{bmatrix} = \begin{bmatrix}\hat{\mu}\\ \hat{h}_1\\ \hat{h}_2\\ \hat{h}_3\\ \hat{g}_1\\ \hat{g}_2\\ \hat{s}_{11}\\ \hat{s}_{12}\\ \hat{s}_{13}\\ \hat{s}_{21}\\ \hat{s}_{22}\end{bmatrix} = \begin{bmatrix}0\\ 13{,}358\\ 14{,}264\\ 13{,}154\\ 0\\ -2{,}056\\ -0{,}048\\ -0{,}038\\ 0{,}087\\ 0{,}389\\ -0{,}389\end{bmatrix}$$

$$E\begin{bmatrix}\hat{\underline{b}}\\ \hat{\underline{u}}\end{bmatrix} = \begin{array}{c} \begin{array}{ccccccccccc} \mu & h_1 & h_2 & h_3 & g_1 & g_2 & s_{11} & s_{12} & s_{13} & s_{21} & s_{22} \end{array} \\ \left[\begin{array}{cccccc|ccccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & & & 1 & & {,}621 & {,}227 & {,}152 & {,}251 & -{,}251 \\ 1 & & 1 & & 1 & & {,}223 & {,}396 & {,}381 & {,}034 & -{,}034 \\ 1 & & & 1 & 1 & & {,}280 & {,}552 & {,}168 & -{,}158 & {,}158 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & & & & -1 & 1 & -{,}306 & -{,}421 & -{,}273 & {,}395 & {,}605 \\ \hline & & & & & & {,}082 & -{,}056 & -{,}026 & -{,}019 & {,}019 \\ & & & & & & -{,}056 & {,}102 & -{,}047 & {,}023 & -{,}023 \\ & & & & & & -{,}026 & -{,}047 & {,}073 & -{,}004 & {,}004 \\ & & & & & & -{,}019 & {,}023 & -{,}004 & {,}121 & -{,}121 \\ & & & & & & {,}019 & -{,}023 & {,}004 & -{,}121 & {,}121 \end{array}\right] \end{array} \begin{bmatrix}\mu\\ h_1\\ h_2\\ h_3\\ g_1\\ g_2\\ s_{11}\\ s_{12}\\ s_{13}\\ s_{21}\\ s_{22}\end{bmatrix}$$

og den inverse koeffisientmatrisen er:

	$\hat{\mu}$	$\hat{h}_1$	$\hat{h}_2$	$\hat{h}_3$	$\hat{g}_1$	$\hat{g}_2$	$\hat{s}_{11}$	$\hat{s}_{12}$	$\hat{s}_{13}$	$\hat{s}_{21}$	$\hat{s}_{22}$
$\hat{\mu}$	0	0	0	0	0	0	0	0	0	0	0
$\hat{h}_1$	0	0,339	0,086	0,067	0	-0,120	-0,041	-0,015	-0,010	-0,017	0,017
$\hat{h}_2$	0		0,216	0,095	0	-0,155	-0,015	-0,026	-0,025	-0,002	0,002
$\hat{h}_3$	0			0,201	0	-0,126	-0,019	-0,037	-0,011	0,011	-0,011
$\hat{g}_1$	0				0	0	0	0	0	0	0
$\hat{g}_2$	0					0,246	0,020	0,028	0,018	-0,026	-0,040
$\hat{s}_{11}$	0	symmetrisk					0,061	0,004	0,002	0,001	-0,001
$\hat{s}_{12}$	0							0,060	0,003	-0,002	0,002
$\hat{s}_{13}$	0								0,062	0,000	-0,000
$\hat{s}_{21}$	0									0,059	0,006
$\hat{s}_{22}$	0										0,059

Om en her beregner Var $(\hat{h}_1 - \hat{h}_2)$ får en:

$$[0{,}339 + 0{,}216 - (2)(0{,}086)]\sigma^2 = \underline{0{,}383\sigma^2}$$

mot $0{,}573\sigma^2$ når både g, $\underline{h}$ og $\underline{s}$ var "fixed" effekter.

Hva er variansen på differansen $(\hat{s}_{11} - s_{11}) - (\hat{s}_{12} - s_{12})$?
(NB! formuleringen når $\underline{s}$ er "random").

$$\text{Var}\,[(\hat{s}_{11} - s_{11}) - (\hat{s}_{12} - s_{12})] = [0{,}061 + 0{,}060 - (2)(0{,}004)]\sigma^2 = \underline{0{,}113\sigma^2}$$

mot $0{,}831\sigma^2$ når både g, $\underline{h}$ og $\underline{s}$ var "fixed" effekter.

Om en önsker å opprettholde en <u>konstant</u> basis, kan "LaGrange multipliers" benyttes. En kan tenke seg fölgende ikke-estimérbare funksjon: $\hat{g}_1 + \hat{s}_{11} + \hat{g}_1 + \hat{s}_{12} = 2\,\hat{g}_1 + \hat{s}_{11} + \hat{s}_{12}$. I tillegg til denne ikke-estimérbare funksjon trenger vi en "constraint" til, f.eks. $\hat{\mu} = 0$.

$$\begin{bmatrix}\hat{\underline{b}}\\ \hline \hat{\underline{u}}\end{bmatrix} = \begin{bmatrix}\hat{\mu}\\ \hat{h}_1\\ \hat{h}_2\\ \hat{h}_3\\ \hat{g}_1\\ \hat{g}_2\\ \hat{s}_{11}\\ \hat{s}_{12}\\ \hat{s}_{13}\\ \hat{s}_{21}\\ \hat{s}_{22}\\ \lambda\end{bmatrix} = \begin{bmatrix}0\\ 13{,}315\\ 14{,}221\\ 13{,}111\\ 0{,}043\\ -2{,}013\\ -0{,}048\\ -0{,}038\\ 0{,}086\\ 0{,}389\\ -0{,}389\\ 0\end{bmatrix}$$

$$E = \begin{bmatrix}\hat{\underline{b}}\\ \hline \hat{\underline{u}}\end{bmatrix} = \begin{array}{c}\begin{array}{ccccccccccc}\mu & h_1 & h_2 & h_3 & g_1 & g_2 & s_{11} & s_{12} & s_{13} & s_{21} & s_{22}\end{array}\\ \left[\begin{array}{cccccc|ccccc}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
1 & 1 & & & & & {,}134 & -{,}250 & {,}115 & {,}253 & -{,}253\\
1 & & 1 & & & & -{,}264 & -{,}081 & {,}345 & {,}036 & -{,}036\\
1 & & & 1 & & & -{,}207 & {,}075 & {,}132 & -{,}156 & {,}156\\
 & & & & 1 & & {,}487 & {,}477 & {,}036 & -{,}002 & {,}002\\
 & & & & & 1 & {,}181 & {,}056 & -{,}237 & {,}393 & {,}607\\ \hline
 & & & & & & {,}082 & -{,}056 & -{,}026 & -{,}019 & {,}019\\
 & & & & & & -{,}056 & {,}102 & -{,}047 & {,}023 & -{,}023\\
 & & & & & & -{,}026 & -{,}047 & {,}073 & -{,}004 & {,}004\\
 & & & & & & -{,}019 & {,}023 & -{,}004 & {,}121 & -{,}121\\
 & & & & & & {,}019 & -{,}023 & {,}004 & -{,}121 & {,}121\\
 & & & & & & & & & &
\end{array}\right]\end{array}
\begin{bmatrix}\mu\\ h_1\\ h_2\\ h_3\\ g_1\\ g_2\\ \hline s_{11}\\ s_{12}\\ s_{13}\\ s_{21}\\ s_{22}\\ \\ \end{bmatrix}$$

sml. korres. tabell med
annen "constraint"

Den inverse koeffisientmatrisen er:

$$
\begin{array}{c|cccccc|ccccc|}
 & \hat{\mu} & \hat{h}_1 & \hat{h}_2 & \hat{h}_3 & \hat{g}_1 & \hat{g}_2 & \hat{s}_{11} & \hat{s}_{12} & \hat{s}_{13} & \hat{s}_{21} & \hat{s}_{22} \\
\hat{\mu} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hat{h}_1 & & 0{,}315 & 0{,}069 & 0{,}043 & -0{,}004 & -0{,}010 & -0{,}009 & 0{,}017 & -0{,}008 & -0{,}017 & 0{,}017 \\
\hat{h}_2 & & & 0{,}207 & 0{,}079 & -0{,}011 & -0{,}142 & 0{,}018 & 0{,}005 & -0{,}023 & -0{,}002 & 0{,}002 \\
\hat{h}_3 & & & & 0{,}178 & -0{,}004 & -0{,}106 & 0{,}014 & -0{,}005 & -0{,}009 & 0{,}010 & -0{,}010 \\
\hat{g}_1 & & & & & 0{,}032 & 0{,}008 & -0{,}032 & -0{,}032 & -0{,}002 & 0{,}000 & -0{,}000 \\
\hat{g}_2 & & & & & & 0{,}230 & -0{,}012 & -0{,}004 & 0{,}016 & -0{,}026 & -0{,}040 \\
\hline
\hat{s}_{11} & \text{symmetrisk} & & & & & & 0{,}061 & 0{,}004 & 0{,}002 & 0{,}001 & -0{,}001 \\
\hat{s}_{12} & & & & & & & & 0{,}060 & 0{,}003 & -0{,}002 & 0{,}002 \\
\hat{s}_{13} & & & & & & & & & 0{,}062 & 0{,}000 & -0{,}000 \\
\hat{s}_{21} & & & & & & & & & & 0{,}059 & 0{,}008 \\
\hat{s}_{22} & & & & & & & & & & & 0{,}059
\end{array}
$$

Hva er Var $(\hat{h}_1 - \hat{h}_2)$?

$$\text{Var } (\hat{h}_1 - \hat{h}_2) = [0{,}315 + 0{,}207 - (2)(0{,}069)]\sigma^2 = \underline{0{,}383\sigma^2}$$

som tidligere.

En bör merke seg at <u>estimater av estimérbare funksjoner av "fixed" effekter og lösningene for "random" effekter blir de samme uansett "constraints". Variansene blir også de samme uansett hvilke ikke-estimérbare funksjoner som er brukt som "constraints"</u>. C_{22} er "unique".

c. <u>$\underline{g}$ er "fixed" effekt men både $\underline{h}$ og $\underline{s}$ er "random" effekter</u>

Vi har da:

$$\underline{y} = X\underline{b} + Z\underline{u} + \underline{e}$$

hvor

$$\underset{\sim}{b} = \begin{bmatrix} \mu \\ g_1 \\ g_2 \end{bmatrix} \quad , \quad \underset{\sim}{u} = \begin{bmatrix} h_1 \\ h_2 \\ h_3 \\ s_{11} \\ s_{12} \\ s_{13} \\ s_{21} \\ s_{22} \end{bmatrix}$$

Nå blir:

$$G^{-1}\frac{1}{\sigma_e^2} = \begin{bmatrix} I_h \dfrac{\sigma_e^2}{\sigma_h^2} & 0 \\ 0 & I_s \dfrac{\sigma_e^2}{\sigma_s^2} \end{bmatrix} \frac{1}{\sigma_e^2}$$

Dette innebärer at en adderer $\frac{\sigma_e^2}{\sigma_h^2}$ til diagonalelementene for besetning og $\frac{\sigma_e^2}{\sigma_s^2}$ til diagonalelementene for fedre. En forutsetter:

$$\frac{\sigma_e^2}{\sigma_h^2} = 2 \; ; \; \frac{\sigma_e^2}{\sigma_s^2} = 15$$

Således:

$$\left[\begin{array}{c|ccc|cc|ccccc}
23 & 4 & 10 & 9 & 9 & 14 & 3 & 4 & 2 & 5 & 9 \\
 & 4+2 & 0 & 0 & 2 & 2 & 2 & 0 & 0 & 2 & 0 \\
 & & 10+2 & 0 & 3 & 7 & 0 & 1 & 2 & 3 & 4 \\
 & & & 9+2 & 4 & 5 & 1 & 3 & 0 & 0 & 5 \\
 & & & & 9 & 0 & 3 & 4 & 2 & 0 & 0 \\
 & & & & & 14 & 0 & 0 & 0 & 5 & 9 \\
 & \text{symmetrisk} & & & & & 3+15 & 0 & 0 & 0 & 0 \\
 & & & & & & & 4+15 & 0 & 0 & 0 \\
 & & & & & & & & 2+15 & 0 & 0 \\
 & & & & & & & & & 5+15 & 0 \\
 & & & & & & & & & & 9+15
\end{array}\right]
\begin{bmatrix} \hat{\mu} \\ \hat{h}_1 \\ \hat{h}_2 \\ \hat{h}_3 \\ \hat{g}_1 \\ \hat{g}_2 \\ \hat{s}_{11} \\ \hat{s}_{12} \\ \hat{s}_{13} \\ \hat{s}_{21} \\ \hat{s}_{22} \end{bmatrix}
=
\begin{bmatrix} 284 \\ 50 \\ 128 \\ 106 \\ 122 \\ 162 \\ 39 \\ 53 \\ 30 \\ 67 \\ 95 \end{bmatrix}$$

Her trenger vi bare én "constraint" da koeffisientene for g summerer opp til koeffisienten for μ. Vi velger å sette $\hat{\mu} = 0$.

En får da:

$$\begin{bmatrix}\underline{b}\\ \hat{\underline{u}}\end{bmatrix} = \begin{bmatrix}\hat{\mu}\\ \hat{h}_1\\ \hat{h}_2\\ \hat{h}_3\\ \hat{g}_1\\ \hat{g}_2\\ \hat{s}_{11}\\ \hat{s}_{12}\\ \hat{s}_{13}\\ \hat{s}_{21}\\ \hat{s}_{22}\end{bmatrix} = \begin{bmatrix}0\\ -0{,}166\\ 0{,}537\\ -0{,}371\\ 13{,}594\\ 11{,}571\\ -0{,}060\\ -0{,}042\\ 0{,}102\\ 0{,}393\\ -0{,}393\end{bmatrix}$$

$$E\begin{bmatrix}\underline{b}\\ \hat{\underline{u}}\end{bmatrix} = \begin{array}{c} \begin{array}{ccccccccccc} \mu & h_1 & h_2 & h_3 & g_1 & g_2 & s_{11} & s_{12} & s_{13} & s_{21} & s_{22} \end{array} \\ \left[\begin{array}{c|ccc|cc|ccccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & {,}460 & -{,}234 & -{,}225 & 0 & 0 & {,}170 & -{,}113 & -{,}057 & {,}143 & -{,}143 \\ 0 & -{,}234 & {,}527 & -{,}293 & 0 & 0 & -{,}110 & -{,}006 & {,}116 & {,}005 & -{,}005 \\ 0 & -{,}225 & -{,}293 & {,}518 & 0 & 0 & -{,}060 & {,}119 & -{,}059 & -{,}148 & {,}148 \\ \hline 1 & {,}299 & {,}342 & {,}359 & 1 & 0 & {,}062 & {,}402 & {,}236 & {,}030 & -{,}030 \\ 1 & {,}280 & {,}375 & {,}345 & 0 & 1 & {,}049 & -{,}019 & -{,}030 & {,}424 & {,}576 \\ 0 & {,}023 & -{,}015 & -{,}008 & 0 & 0 & {,}091 & -{,}061 & -{,}030 & -{,}013 & {,}013 \\ 0 & -{,}015 & -{,}001 & {,}016 & 0 & 0 & -{,}061 & {,}108 & -{,}047 & {,}017 & -{,}017 \\ 0 & -{,}008 & {,}015 & -{,}008 & 0 & 0 & -{,}030 & -{,}047 & {,}076 & -{,}004 & {,}004 \\ 0 & {,}019 & {,}001 & -{,}020 & 0 & 0 & -{,}013 & {,}017 & -{,}004 & {,}129 & -{,}129 \\ 0 & -{,}019 & -{,}001 & {,}020 & 0 & 0 & {,}013 & -{,}017 & {,}004 & -{,}129 & {,}129 \end{array}\right] \end{array} \begin{bmatrix}\mu\\ h_1\\ h_2\\ h_3\\ g_1\\ g_2\\ s_{11}\\ s_{12}\\ s_{13}\\ s_{21}\\ s_{22}\end{bmatrix}$$

og den inverse koeffisientmatrisen:

	$\hat{\mu}$	$\hat{h}_1$	$\hat{h}_2$	$\hat{h}_3$	$\hat{g}_1$	$\hat{g}_2$	$\hat{s}_{11}$	$\hat{s}_{12}$	$\hat{s}_{13}$	$\hat{s}_{21}$	$\hat{s}_{22}$
$\hat{\mu}$	0	0	0	0	0	0	0	0	0	0	0
$\hat{h}_1$		0,270	0,117	0,113	-0,150	-0,140	-0,011	0,008	0,004	-0,010	0,010
$\hat{h}_2$			0,237	0,146	-0,171	-0,187	0,007	0,000	-0,008	-0,000	0,000
$\hat{h}_3$				0,241	-0,180	-0,172	0,004	-0,008	0,004	0,010	-0,010
$\hat{g}_1$					0,305	0,170	-0,024	-0,027	-0,016	-0,002	0,002
$\hat{g}_2$						0,282	-0,003	0,001	0,002	-0,028	-0,038
$\hat{s}_{11}$	symmetrisk						0,061	0,004	0,002	0,001	-0,001
$\hat{s}_{12}$								0,060	0,003	-0,001	0,001
$\hat{s}_{13}$									0,062	0,000	-0,000
$\hat{s}_{21}$										0,058	0,009
$\hat{s}_{22}$											0,058

Hva er variansen på differansen $(\hat{h}_1 - \hat{h}_1) - (\hat{h}_2 - \hat{h}_2)$ her ?

$\text{Var}[(\hat{h}_1 - h_1) - (\hat{h}_2 - h_2)] = [0{,}270 + 0{,}237 - (2)(0{,}117)]\sigma^2 =$
$\underline{0{,}273\sigma^2}$

Som et eksempel på hvordan variansen avhenger av om effekten er "fixed" eller "random", skal vi for $(\hat{h}_1 - \hat{h}_2)$ [eller $(\hat{h}_1 - \hat{h}_1) - (\hat{h}_2 - \hat{h}_2)$] og $(\hat{s}_{11} - \hat{s}_{12})$ [eller $(\hat{s}_{11} - s_{11}) - (\hat{s}_{12} - s_{12})$] sammenstille resultatene:

Modell (f = "fixed" ; r = "random"):

$\underline{h}$	g	$\underline{s}$	Var. av	Var. av
f	f	f	$(\hat{h}_1 - \hat{h}_2) = 0{,}573\sigma^2$	$(\hat{s}_{11} - \hat{s}_{12}) = 0{,}830\sigma^2$
f	f	r	$(\hat{h}_1 - \hat{h}_2) = 0{,}383\sigma^2$	$(\hat{s}_{11}-s_{11})-(\hat{s}_{12}-s_{12}) = 0{,}113\sigma^2$
r	f	r	$(\hat{h}_1-h_1)-(\hat{h}_2-h_2) = 0{,}273\sigma^2$	$(\hat{s}_{11}-s_{11})-(\hat{s}_{12}-s_{12}) = 0{,}111\sigma^2$

Om besetningseffekten egentlig er "random", men vi bruker den som "fixed", <u>blir variansen på estimatene av estimérbare funksjoner av "fixed" effekter og variansen på "predictions" större enn om besetning hadde blitt behandlet som "random" effekt i analysen.</u> Estimatene og "predictions" er imidlertid "unbiased" (se IX).

d. <u>$\underline{h}$, g er "fixed" effekter mens $\underline{s}$ er "random" effekter. Fedrene er beslektet (se også V D)</u>.

La oss forutsette at fedrene er beslektet som vist nedenfor:

$$A = \begin{array}{c} \\ s_{11} \\ s_{12} \\ s_{13} \\ s_{21} \\ s_{22} \end{array} \begin{array}{c} \begin{array}{ccccc} s_{11} & s_{12} & s_{13} & s_{21} & s_{22} \end{array} \\ \begin{bmatrix} 1 & 1/4 & 1/2 & 0 & 0 \\ 1/4 & 1 & 1/8 & 1/4 & 1/4 \\ 1/2 & 1/8 & 1 & 0 & 0 \\ 0 & 1/4 & 0 & 1 & 1/2 \\ 0 & 1/4 & 0 & 1/2 & 1 \end{bmatrix} \end{array}$$

Dette betyr at f.eks. s_{11} og s_{12} er like mye beslektet som halvsösken. Videre er ingen av fedrene innavlet.

Fölgelig er $G^{-1}\frac{1}{\sigma_e^2} = \left[A^{-1}\ \frac{\sigma_e^2}{\sigma_s^2}\right] \frac{1}{\sigma_e^2}$, og G^{-1} skal adderes til $Z'Z$ i BLUP-ligningene.

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z+G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\underline{b}} \\ \hat{\underline{u}} \end{bmatrix} = \begin{bmatrix} X'\underline{y} \\ Z'\underline{y} \end{bmatrix}$$

eller,

$$\left[\begin{array}{cccccc|ccccc} 23 & 4 & 10 & 9 & 9 & 14 & 3 & 4 & 2 & 5 & 9 \\ & 4 & 0 & 0 & 2 & 2 & 2 & 0 & 0 & 2 & 0 \\ & & 10 & 0 & 3 & 7 & 0 & 1 & 2 & 3 & 4 \\ & & & 9 & 4 & 5 & 1 & 3 & 0 & 0 & 5 \\ & & & & 9 & 0 & 3 & 4 & 2 & 0 & 0 \\ & & & & & 14 & 0 & 0 & 0 & 5 & 9 \\ \hline & & & & & & 3+ & 0+ & 0+ & 0+ & 0+ \\ \text{symmetrisk} & & & & & & 21{,}10 & -4{,}39 & -10{,}00 & {,}73 & {,}73 \\ & & & & & & & 4+ & & 0+ & 0+ \\ & & & & & & & 17{,}56 & 0 & -2{,}93 & -2{,}93 \\ & & & & & & & & 2+ & & \\ & & & & & & & & 20 & 0 & 0 \\ & & & & & & & & & 5+ & 0+ \\ & & & & & & & & & 20{,}49 & -9{,}51 \\ & & & & & & & & & & 9+ \\ & & & & & & & & & & 20{,}49 \end{array}\right] \left[\begin{array}{c} \hat{\mu} \\ \hat{h}_1 \\ \hat{h}_2 \\ \hat{h}_3 \\ \hat{g}_1 \\ \hat{g}_2 \\ \hline \hat{s}_{11} \\ \hat{s}_{12} \\ \hat{s}_{13} \\ \hat{s}_{21} \\ \hat{s}_{22} \end{array}\right] = \left[\begin{array}{c} 284 \\ 50 \\ 128 \\ 106 \\ 122 \\ 162 \\ \hline 39 \\ 53 \\ 30 \\ 67 \\ 95 \end{array}\right]$$

De nederste sifrene i $Z'Z+G^{-1}$ er G^{-1}. $(\frac{\sigma_e^2}{\sigma_s^2} = 15)$.

Vi må innföre to "constraints", og velger å bruke:

$$2\ \hat{g}_1 + \hat{s}_{11} + \hat{s}_{12} = 0$$

$$\hat{\mu} = 0.$$

Dette gir:

$$
\begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} \hat{\mu} \\ \hat{h}_1 \\ \hat{h}_2 \\ \hat{h}_3 \\ \hat{g}_1 \\ \hat{g}_2 \\ \hat{s}_{11} \\ \hat{s}_{12} \\ \hat{s}_{13} \\ \hat{s}_{21} \\ \hat{s}_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 13{,}425 \\ 14{,}258 \\ 13{,}054 \\ 0{,}028 \\ -2{,}068 \\ -0{,}024 \\ -0{,}031 \\ 0{,}054 \\ 0{,}215 \\ -0{,}228 \end{bmatrix}
\qquad
E\begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} =
\begin{array}{c}
\begin{array}{ccccccccccc} \mu & h_1 & h_2 & h_3 & g_1 & g_2 & s_{11} & s_{12} & s_{13} & s_{21} & s_{22} \end{array} \\
\left[\begin{array}{cccccc|ccccc}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & & & & & ,133 & -,297 & ,114 & ,287 & -,287 \\
1 & 0 & 1 & & & & -,284 & -,075 & ,359 & ,044 & -,044 \\
1 & 0 & 0 & 1 & & & -,216 & ,076 & ,140 & -,174 & ,175 \\
0 & & & & 1 & 0 & ,491 & ,485 & ,023 & -,001 & ,001 \\
0 & & & & 0 & 1 & ,211 & ,023 & -,235 & ,369 & ,631 \\
\hline
 & & & & & & ,058 & -,055 & -,003 & -,018 & ,018 \\
 & & & & & & -,140 & ,084 & -,044 & ,021 & -,021 \\
 & & & & & & ,070 & -,062 & ,055 & -,012 & ,012 \\
 & & & & & & -,026 & ,040 & -,014 & ,076 & -,076 \\
 & & & & & & -,003 & ,013 & -,010 & -,062 & ,062
\end{array}\right]
\end{array}
\begin{bmatrix} \mu \\ h_1 \\ h_2 \\ h_3 \\ g_1 \\ g_2 \\ \hline s_{11} \\ s_{12} \\ s_{13} \\ s_{21} \\ s_{22} \end{bmatrix}
$$

og den inverse koeffisientmatrisen er:

	$\hat{\mu}$	$\hat{h}_1$	$\hat{h}_2$	$\hat{h}_3$	$\hat{g}_1$	$\hat{g}_2$	$\hat{s}_{11}$	$\hat{s}_{12}$	$\hat{s}_{13}$	$\hat{s}_{21}$	$\hat{s}_{22}$
$\hat{\mu}$	0	0	0	0	0	0	0	0	0	0	0
$\hat{h}_1$		0,309	0,069	0,047	-0,002	-0,102	-0,009	0,013	-0,010	-0,005	0,014
$\hat{h}_2$			0,201	0,076	-0,007	-0,139	0,008	0,007	-0,014	-0,000	0,003
$\hat{h}_3$				0,174	-0,003	-0,104	0,008	-0,003	-0,003	0,005	-0,007
$\hat{g}_1$					0,041	0,013	-0,042	-0,041	-0,022	-0,008	-0,008
$\hat{g}_2$						0,244	-0,007	-0,020	0,008	-0,046	-0,055
$\hat{s}_{11}$	symmetrisk						0,064	0,019	0,032	0,002	0,000
$\hat{s}_{12}$								0,062	0,012	0,015	0,016
$\hat{s}_{13}$									0,063	0,001	0,001
$\hat{s}_{21}$										0,063	0,035
$\hat{s}_{22}$											0,064

Hva er Var $(\hat{h}_1 - \hat{h}_2)$ her ?

$$\text{Var } (\hat{h}_1 - \hat{h}_2) = [0{,}309 + 0{,}201 - (2)(0{,}069)]\sigma^2 = \underline{0{,}372\sigma^2}$$

Videre:

$$\text{Var } [(\hat{s}_{11} - s_{11}) - (\hat{s}_{12} - s_{12})] = [0{,}064 + 0{,}062 - (2)(0{,}019)]\sigma^2 = \underline{0{,}088\sigma^2}$$

Ved å ta inn slektskapet mellom individene får en således mindre varians både på estimatene av estimérbare funksjoner av "fixed" effekter og på "predictions" (sml. tidligere resultat s. V A:11).

EKSEMPEL B. RANGERING AV OKSER PÅ GRUNNLAG AV ALDERSKORRIGERTE DATA FOR MJÖLKEAVKASTNING (SCHAEFFER, 1975)

La oss arbeide med fölgende data:

Besetning-år-sesong	Identifikasjon av okser	Antall avkom	Totalt for avkom $\frac{\text{Mjölkemengde}}{100}$
1	20021	4	531
1	20032	3	449
1	20103	3	416
2	20032	3	411
2	20114	2	298
3	20103	4	624
3	20235	6	983
4	20021	2	302
4	20266	3	526
5	20021	2	321
6	20021	2	254
6	20114	4	746
6	20266	2	363

Totalt har vi 40 observasjoner, 6 grupper av besetning-år-sesong kombinasjon og 6 okser. Den underliggende modellen her er:

$$y_{ijkl} = g_i + s_{ij} + h_k + e_{ijkl} ,$$

hvor

y_{ijkl} = mjölkeavkastningen til l^{te} datter etter j^{te} far i i^{te} gruppe og k^{te} besetning-år-sesong kombinasjon

g_i = konstant som er felles for alle dötre etter fedre i i^{te} gruppe ("fixed" effekt)

s_{ij} = effekten av j^{te} far i i^{te} gruppe på sine dötre ("random" effekt)

h_k = effekten av k besetning-år-sesong kombinasjon ("fixed" effekt)

e_{ijkl} = tilfeldig effekt på hver registrering

E(s, e) = 0, og hanndyrene forutsettes å väre ubeslektet.

Selve grupperingen av fedre går ut på å definere hvilken populasjon som de ble samplet fra. Her kan en velge mellom flere kriterier f.eks. födselsår, testingsår, avstamningsresultat osv. Videre bör en ha 20-40 fedre pr. gruppe. I det fölgende bruker vi 2 grupper, nemlig:

$$g_1 \begin{cases} s_{11} = 20021 \\ s_{12} = 20103 \\ s_{13} = 20114 \end{cases} \qquad g_2 \begin{cases} s_{21} = 20032 \\ s_{22} = 20235 \\ s_{23} = 20266 \end{cases}$$

"Mixed-model" ligningene er gitt nedenfor:

$$\begin{array}{c} \begin{array}{cccccccccccccc} g_1 & g_2 & s_{11} & s_{21} & s_{12} & s_{13} & s_{22} & s_{23} & h_1 & h_2 & h_3 & h_4 & h_5 & h_6 \end{array} \\ \begin{bmatrix} 23 & 0 & 10 & 0 & 7 & 6 & 0 & 0 & 7 & 2 & 4 & 2 & 2 & 6 \\ & 17 & 0 & 6 & 0 & 0 & 6 & 5 & 3 & 3 & 6 & 3 & 0 & 2 \\ & & 25^* & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 2 & 2 & 2 \\ & & & 21 & 0 & 0 & 0 & 0 & 3 & 3 & 0 & 0 & 0 & 0 \\ & & & & 22 & 0 & 0 & 0 & 3 & 0 & 4 & 0 & 0 & 0 \\ & & & & & 21 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 4 \\ & & & & & & 21 & 0 & 0 & 0 & 6 & 0 & 0 & 0 \\ & & & & & & & 20 & 0 & 0 & 0 & 3 & 0 & 2 \\ & & & & & & & & 10 & 0 & 0 & 0 & 0 & 0 \\ & & & & & & & & & 5 & 0 & 0 & 0 & 0 \\ \text{symmetrisk} & & & & & & & & & & 10 & 0 & 0 & 0 \\ & & & & & & & & & & & 5 & 0 & 0 \\ & & & & & & & & & & & & 2 & 0 \\ & & & & & & & & & & & & & 8 \end{bmatrix} \end{array} \begin{bmatrix} \hat{g}_1 \\ \hat{g}_2 \\ \hat{s}_{11} \\ \hat{s}_{21} \\ \hat{s}_{12} \\ \hat{s}_{13} \\ \hat{s}_{22} \\ \hat{s}_{23} \\ \hat{h}_1 \\ \hat{h}_2 \\ \hat{h}_3 \\ \hat{h}_4 \\ \hat{h}_5 \\ \hat{h}_6 \end{bmatrix} = \begin{bmatrix} 3492 \\ 2732 \\ 1408 \\ 860 \\ 1040 \\ 1044 \\ 983 \\ 889 \\ 1396 \\ 709 \\ 1607 \\ 828 \\ 321 \\ 1363 \end{bmatrix}$$

$^*10 + 15 = 25 \qquad G^{-1}$

Hvordan en teknisk sett dannet <u>gruppeligningene</u> kan kanskje väre nyttig å illustrere mer detaljert. Vi hadde altså 6 okser, og ut fra et kriterium (eller flere kriterier) bestemte en at gruppe 1 skulle omfatte oksene s_{11}, s_{12} og s_{13} mens gruppe 2 skulle omfatte oksene s_{21}, s_{22} og s_{23}. Dette innebärer at f.eks.:

$\hat{s}_{11}$ i g_1-ligningen er: $10 + 0 + 0 = \underline{10}$

$\hat{g}_2$ i g_1-ligningen (og $\hat{g}_1$ i g_2-ligningen) er: $0 + 0 + 0 = 0$
dvs. ($s_{21} + s_{22} + s_{23}$ i g_1)., osv.

Ved å innföre 1 "constraint", eks. $\hat{g}_1 = 0$, får en fölgende lösninger:

$\hat{s}_{11} = -5{,}39291$	$\hat{s}_{21} = -0{,}91929$
$\hat{s}_{12} = 0{,}15664$	$\hat{s}_{22} = -0{,}25953$
$\hat{s}_{13} = 5{,}23627$	$\hat{s}_{23} = 1{,}17882$
$\hat{g}_1 = 0$	$\hat{g}_2 = 9{,}87155$
$\hat{h}_1 = 139{,}02450$	$\hat{h}_2 = 134{,}33414$
$\hat{h}_3 = 154{,}87014$	$\hat{h}_4 = 161{,}12694$
$\hat{h}_5 = 165{,}89291$	$\hat{h}_6 = 166{,}34250$

En skal merke seg at uansett "constraints" blir:

- $\hat{g}_1 - \hat{g}_2 = -9{,}87155$
- forskjellene mellom $\hat{h}$ konstant
- lösningene for $\hat{s}$ de samme. En ser også at $\Sigma s_j = 0$ fordi varians-kovarians matrisen (G) er diagonal, og alle diagonalelementene er lik samme konstant. Om <u>diagonalene til G ikke er lik, eller om "off"-diagonal-elementene eksisterer i G, eller $R \neq I\sigma_e^2$, summerer "random" effektene nödvendigvis ikke til 0</u>.

Av resultatene ovenfor kan en se at far 1 i gruppe 1 får verdien ("sire proof"): $0{,}0 - 5{,}39291 = -5{,}39291$ (NB. vär konsekvent ved valg av basis).

Når det gjelder mormaterialet som hanndyrene pares til, må en i disse beregningene forutsette at det i gjennomsnitt er det samme for alle fedre som testes. Holder ikke denne forutsetningen, må en som påpekt av Henderson (1973), ta både morfedre og fedre inn i modellen (se V J:1).

EKSEMPEL C. RANGERING AV OKSER NÅR EN BRUKER FLERE LAKTASJONER PR. DATTER (SCHAEFFER, 1975)

La oss forutsette at vi skal arbeide med data som er vist i tabellen nedenfor for rangering av okser:

		Registrering					
Far	Datter	Besetning-år-sesong kombinasjon	1.*	Besetning-år-sesong kombinasjon	2.	Besetning-år-sesong kombinasjon	3.
20021	1	1	104	3	110	-	-
20021	2	1	103	3	122	-	-
20021	3	1	179	3	142	5	137
20021	4	1	145	3	177	5	199
20021	26	4	139	6	105	-	-
20021	27	4	163	6	191	8	200
20021	31	5	170	7	117	-	-
20021	32	5	151	7	146	9	192
20021	37	6	113	8	197	10	201
20021	38	6	141	8	144	10	142
20032	5	1	156	3	196	5	189
20032	6	1	150	3	185	-	-
20032	7	1	143	3	128	-	-
20032	13	2	166	-	-	-	-
20032	14	2	102	-	-	-	-
20032	15	2	143	4	102	-	-
20103	8	1	127	-	-	-	-
20103	9	1	185	3	163	-	-
20103	10	1	104	-	-	-	-
20103	16	3	129	5	101	-	-
20103	17	3	188	5	152	-	-
20103	18	3	174	-	-	-	-
20103	19	3	133	5	148	7	154
20114	11	2	110	-	-	-	-
20114	12	2	188	-	-	-	-
20114	33	6	186	8	134	10	170
20114	34	6	186	8	198	-	-
20114	35	6	185	8	191	10	140
20114	36	6	189	8	163	-	-
20235	20	3	197	5	132	-	-
20235	21	3	189	5	129	-	-
20235	22	3	148	5	102	-	-
20235	23	3	105	-	-	-	-
20235	24	3	163	5	181	7	130
20235	25	3	181	5	100	-	-
20266	28	4	198	6	201	-	-
20266	29	4	145	6	169	8	126
20266	30	4	183	6	191	-	-
20266	39	6	187	8	193	-	-
20266	40	6	176	8	175	10	166

*1. laktasjon

En går ut fra at det er foretatt en riktig korrigering for effekten av alder og kalvingstid.

Den modellen som ligger til grunn for disse data, kan sies å väre:

Registrering = effekten av fargruppe + effekten av fedre innen gruppe + effekten av besetning-år-sesong + effekten av kyr innen fedre og besetning + rest.

Uttrykt ved "mixed model" får en når en for enkelhets skyld forutsetter bare <u>en</u> besetning:

$$\underline{y} = X\underline{g} + Z\underline{s} + W\underline{h} + P\underline{c} + \underline{e} \text{ , hvor}$$

$\underline{y}$ = vektor av observasjoner

$\underline{g}$ = vektor av far-gruppe-effekter ("fixed")

$\underline{s}$ = vektor av far-effekter ("random")

$\underline{h}$ = vektor av besetning-år-sesong-effekter ("fixed")

$\underline{c}$ = vektor av ku-effekter ("random")

$\underline{e}$ = vektor av tilfeldige feil

Nå blir $\sigma_s^2 = \frac{1}{4}h^2\sigma_y^2$; $\sigma_c^2 = (r - \frac{1}{4}h^2)\sigma_y^2$; $\sigma_e^2 = (1 - r)\sigma_y^2$ hvor

σ_s^2 = varianskomponenten mellom fedre

σ_c^2 = varianskomponenten mellom kyr

σ_e^2 = varianskomponenten for rest

r = reproduserbarheten

Fölgelig må G-matrisen bli:

$$G = \begin{bmatrix} I\sigma_s^2 & 0 \\ 0 & I\sigma_c^2 \end{bmatrix}$$

Videre må en forutsette at alle kyr har <u>förste laktasjonsavkastning registrert</u>, og at alle laktasjoner på samme ku er registrert i <u>samme</u> besetning.

Når det gjelder forholdet mellom varianskomponentene, går en ut fra at:

$$\frac{\sigma_e^2}{\sigma_c^2} = \frac{1 - r}{r - \frac{1}{4}h^2} = 1,0 \quad \text{og} \quad \frac{\sigma_e^2}{\sigma_s^2} = 7,5 \quad (h^2 = 0,25\ ;\ r = 0,53)$$

$E\ (\underline{y}) = Xg + W\underline{h}$

$\text{Var}\ (\underline{y}) = ZZ'\sigma_s^2 + PP'\sigma_c^2 + I\sigma_e^2$, og "mixed model"-ligningene:

$$\begin{bmatrix} X'X & X'Z & X'W & X'P \\ Z'X & Z'Z+I(7,5) & Z'W & Z'P \\ W'X & W'Z & W'W & W'P \\ P'X & P'Z & P'W & P'P+I(1) \end{bmatrix} \begin{bmatrix} \hat{\underline{g}} \\ \hat{\underline{s}} \\ \hat{\underline{h}} \\ \hat{\underline{c}} \end{bmatrix} = \begin{bmatrix} X'\underline{y} \\ Z'\underline{y} \\ W'\underline{y} \\ P'\underline{y} \end{bmatrix}$$

hvor g_1 inkluderer far nr. 20021, 20103 og 20114,
og g_2 inkluderer far nr. 20032, 20235 og 20266.

Dimensjonen på denne matrisen er lik 58 (dvs. 2 far-grupper; 6 fedre; 40 kyr og 10 besetninger).

En vil i det fölgende bare vise noen av ligningene:

$$X'X = \begin{bmatrix} 50 & 0 \\ 0 & 35 \end{bmatrix}$$

$$X'Z = \begin{bmatrix} 26 & 0 & 12 & 12 & 0 & 0 \\ 0 & 11 & 0 & 0 & 12 & 12 \end{bmatrix}$$

$$X'W = \begin{bmatrix} 7 & 2 & 9 & 2 & 7 & 8 & 3 & 7 & 1 & 4 \\ 3 & 3 & 9 & 4 & 6 & 5 & 1 & 3 & 0 & 1 \end{bmatrix}$$

Diagonalene av Z'Z + I (7,5) er :

(33,5 18,5 19,5 19,5 19,5 19,5)

Elementene av Z'W er:

$$\begin{bmatrix} 4 & 0 & 4 & 2 & 4 & 4 & 2 & 3 & 1 & 2 \\ 3 & 3 & 3 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 3 & 0 & 5 & 0 & 3 & 0 & 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 4 & 0 & 4 & 0 & 2 \\ 0 & 0 & 6 & 0 & 5 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 5 & 0 & 3 & 0 & 1 \end{bmatrix}$$

Diagonalene av W'W er:

(10 5 18 6 13 13 4 10 1 5)

Höyresiden (HS) av ligningene er:

Ligning	HS	Ligning	HS	Ligning	HS	Ligning	HS
g_1	7728	c_1	214	c_{16}	230	c_{31}	287
g_2	5527	c_2	225	c_{17}	340	c_{32}	489
		c_3	458	c_{18}	174	c_{33}	490
s_1	3930	c_4	521	c_{19}	435	c_{34}	384
s_2	1660	c_5	541	c_{20}	329	c_{35}	516
s_3	1758	c_6	335	c_{21}	318	c_{36}	352
s_4	2040	c_7	271	c_{22}	250	c_{37}	511
s_5	1757	c_8	127	c_{23}	105	c_{38}	427
s_6	2110	c_9	348	c_{24}	474	c_{39}	380
		c_{10}	104	c_{25}	281	c_{40}	517
h_1	1396	c_{11}	110	c_{26}	244		
h_2	709	c_{12}	188	c_{27}	554		
h_3	2830	c_{13}	166	c_{28}	399		
h_4	930	c_{14}	102	c_{29}	440		
h_5	1891	c_{15}	245	c_{30}	374		
h_6	2220						
h_7	547						
h_8	1721						
h_9	192						
h_{10}	819						

På grunn av kravet på estimérbarhet må en innföre en "constraint", eksempelvis $\hat{g}_1 = 0$. I neste tabell viser en lösningene og de inverse diagonal-elementene.

Lösning til	Lösning	Diagonal*	Lösning til	Lösning	Diagonal*
g_1	0	0	c_{12}	19,7128	0,6190
g_2	3,5747	0,2519	c_{13}	8,7984	0,6096
s_1	-2,1980	0,0959	c_{14}	-23,2016	0,6096
s_2	-1,1479	0,1077	c_{15}	-19,3807	0,4492
s_3	-0,3999	0,1052	c_{16}	-22,0645	0,4080
s_4	2,5980	0,1033	c_{17}	14,6021	0,4080
s_5	-2,7566	0,1094	c_{18}	8,8104	0,5469
s_6	3,9045	0,1127	c_{19}	2,2942	0,3548
h_1	139,9114	0,2476	c_{20}	10,1234	0,4090
h_2	145,9764	0,4995	c_{21}	6,4568	0,4090
h_3	156,7792	0,1947	c_{22}	-16,2099	0,4090
h_4	152,3122	0,3767	c_{23}	-26,2987	0,5484
h_5	140,2143	0,2263	c_{24}	11,1307	0,3599
h_6	165,2035	0,2372	c_{25}	-5,8766	0,4090
h_7	130,0296	0,4645	c_{26}	-23,0399	0,4283
h_8	167,4384	0,2654	c_{27}	18,9100	0,3527
h_9	183,8139	1,4802	c_{28}	22,1753	0,4292
h_{10}	164,6620	0,3963	c_{29}	-15,8479	0,3596
c_1	-26,0982	0,4040	c_{30}	13,8420	0,4292
c_2	-22,4315	0,4040	c_{31}	7,0507	0,4320
c_3	6,9223	0,3294	c_{32}	10,3841	0,4320
c_4	22,6723	0,3294	c_{33}	-3,7745	0,3625
c_5	24,2037	0,3617	c_{34}	15,3874	0,4217
c_6	11,1519	0,4281	c_{35}	2,7255	0,3625
c_7	-10,1814	0,4281	c_{36}	4,7207	0,4217
c_8	-6,2557	0,5595	c_{37}	5,0725	0,3585
c_9	17,3697	0,4064	c_{38}	-15,9275	0,3585
c_{10}	-17,7557	0,5595	c_{39}	10,7999	0,4300
c_{11}	-19,2872	0,6190	c_{40}	-0,6854	0,3779

*for den inverse matrisen.

Resultatet ("sire proof") for det enkelte hanndyr beregnes som tidligere og er vist nedenfor:

Far	Lösning far-effekt	Lösning gruppe-effekt	"Sire proof"
s_1 (20021)	-2,1980	0	-2,1980
s_2 (20032)	-1,1479	3,5747	2,4268
s_3 (20103)	-0,3999	0	-0,3999
s_4 (20114)	2,5980	0	2,5980
s_5 (20235)	-2,7566	3,5747	0,8181
s_6 (20266)	3,9045	3,5747	7,4792

Lösningen for den enkelte fareffekt kan selvsagt også finnes ut fra lösningene for kueffektene (dattereffektene). F.eks.:

$$s_6 = \frac{c_{28} + c_{29} + c_{30} + c_{39} + c_{40}}{7,5} = 3,9045$$

Et interessant spörsmål er om vi også kan beregne avlsverdien til kyrne. Her kan en skille på to definisjoner, nemlig (NB! endret i forhold til Schaeffer, 1975):

a. ETA, "estimated transmitting ability" (Hele genotypen) =

$$g_i + s_{ij} + \underbrace{\sqrt{\frac{\frac{3}{4}h^2}{r - \frac{1}{4}h^2}}}_{\text{fraksjonen av variasjonen i c som er genetisk betinget.}} c_{ijk}$$

Her er den permanente miljöeffekten eliminert, og estimatet gir en nyttig informasjon når en skal avgjöre om flere kalver skal settes på etter den enkelte ku.

b. RPA, "estimated real producing ability" (= "prediction of a future record") = $g_i + s_{ij} + c_{ijk}$. RPA inneholder også eventuell permanent miljöeffekt, og RPA gir en nyttig informasjon når det skal avgjöres om en skal beholde vedkommende ku i besetningen.

En kan dermed sette opp fölgende for f.eks. ku nr. 5:

$$ETA = \hat{g}_i + \hat{s}_{ij} + \sqrt{\frac{\frac{3}{4}h^2}{r - \frac{1}{4}h^2}}\, \hat{c}_{ijk} = 3,5747 - 1,1479 + (0,63)(24,2037) =$$

$$= \underline{17,6751}$$

Dette estimatet er fri for "bias" på grunn av seleksjon og genetisk trend. Derimot er det en viktig forutsetning at det ikke er en genetisk variasjon mellom besetninger. Videre:

$$RPA = \hat{g}_i + \hat{s}_{ij} + \hat{c}_{ijk} = 3,5747 - 1,1479 + 24,2037 = \underline{26,6305}$$

Når det gjelder selve beregningen teknisk sett, er det flere måter å gjöre dette på (Henderson, 1975d). En måte som ofte blir brukt, når det er avlsverdien til hanndyrene (faren) en först og fremst har interesse av å estimere, er fölgende (for enkelhets skyld ser vi bort fra g):

a. Sortêr kyrne innen besetning. Når en har sortert ferdig for den enkelte ku, skulle fölgende informasjon väre tilgjengelig:

- n = antall registreringer ("records") på probanden
- y_i = den enkelte registrering (i = 1 ... n)
- år-sesong for den enkelte registrering
- informasjon om faren til kua.

b. En absorberer så den enkelte c inn i ligningene for besetning-år-sesong og far. Dette gjöres enklest ved å addere for den enkelte besetning fölgende faktorer:

• $\frac{n\alpha}{n+\alpha}$ til diagonalkoeffisienten for faren til kua

$\alpha = [\frac{1-r}{r-0{,}25\ h^2}]$

•• $\frac{\alpha}{n+\alpha}$ til <u>hver</u> av koeffisientene for far x år-sesong som kua hadde sin produksjon i dvs. n adderinger

••• $\frac{n-1+\alpha}{n+\alpha}$ til de n diagonalkoeffisientene for år-sesong

•••• $\frac{-1}{n+\alpha}$ til "off"-diagonalelementene for år-sesong x år sesong

••••• $\frac{\alpha\sum_i y_i}{n+\alpha}$ til höyresiden av ligningen for vedkommende far

•••••• $y_k - \frac{\sum_i y_i}{n+\alpha}$ til höyresiden for k^{te} år-sesong

c. Da absorbsjonen av $\underline{c}$ er ferdig, kan BLUP-ligningene skrives som:

$$\begin{bmatrix} Q_{11} & Q_{12} \\ Q_{12}' & Q_{22} \end{bmatrix} \begin{bmatrix} \hat{\underline{s}} \\ \hat{\underline{h}} \end{bmatrix} = \begin{bmatrix} \underline{f}_1 \\ \underline{f}_2 \end{bmatrix}$$

Det neste blir da å absorbere $\underline{h}$, og en får: $K\hat{\underline{s}} = \underline{d}$ hvor

$K = Q_{11} - Q_{12}Q_{22}^{-1}Q_{12}'$ og

$\underline{d} = \underline{f}_1 - Q_{12}Q_{22}^{-1}\underline{f}_2$

$\underline{s}$ kan da estimeres uten större vanskeligheter.

En raskere metode for absorbsjon av $\underline{h}$ er kanskje fölgende:

$$\begin{bmatrix} Q_{11} & Q_{12} \\ Q_{12}' & Q_{22} \end{bmatrix} \begin{bmatrix} \hat{\underline{s}} \\ \hat{\underline{h}} \end{bmatrix} = \begin{bmatrix} \underline{f}_1 \\ \underline{f}_2 \end{bmatrix}$$ skrives som:

$B\underline{x} = \underline{m}$ $[B_{q\times q}]$

Vi eliminerer förts x_q, og den nye B-matrisen (B_{q-1}) får da fölgende element:

$$b'_{ij} = b_{ij} - \frac{b_{iq}\, b_{jq}}{b_{qq}}$$

$$m'_i = m_i - \frac{b_{iq}\, m_q}{b_{qq}}$$

osv. inntil en har ligningene for $\hat{\underline{s}}$ tilbake.

EKSEMPEL D. RANGERING AV OKSER PÅ GRUNNLAG AV ALDERSKORRIGERTE DATA FOR MJÖLKEAVKASTNING NÅR EN KJENNER KOVARIANSEN (ADDITIVE GENETISKE SLEKTSKAPET) MELLOM OKSENE

1. Schaeffer (1975)

Vi går ut fra samme data som i eksempel B, og forutsetter at vi kjenner avstamningen på samtlige okser. Som basisgenerasjon velger vi en generasjon hvor foreldrene er ubeslektet, eller svärt lite beslektet.

Modellen kan nå skrives på fölgende måte (NB! gruppe-effekten kan da ignoreres):

$$y_{jkl} = s_j + h_k + e_{jkl} \text{ , og}$$

$R = I\sigma_e^2$, $G = A\sigma_s^2$, hvor A inneholder koeffisientene for slektskap mellom okser. For å summere opp får en ved å bruke A beregnet far-effektene med större presisjon, en trenger färre grupper for estimering av genetisk trend og genetisk forskjell mellom populasjoner samtidig som en kan få et relativt sikkert estimat på avlsverdien til hanndyrene tidlig i deres liv.

Om $\frac{\sigma_e^2}{\sigma_s^2} = 15 = r^*$, blir $G^{-1} = A^{-1} r^*$

Videre:

$20021 = s_1$
$20103 = s_2$
$20114 = s_3$
$20032 = s_4$
$20235 = s_5$
$20266 = s_6$

og (bruk kovarians-metoden)

$$A = \begin{array}{c|cccccc|} & s_1 & s_2 & s_3 & s_4 & s_5 & s_6 \\ s_1 & 1 & 0 & 0 & 0 & 0 & 0 \\ s_2 & 0 & 1 & 1/2 & 3/4 & 3/4 & 3/4 \\ s_3 & 0 & 1/2 & 1 & 3/4 & 3/4 & 3/4 \\ s_4 & 0 & 3/4 & 3/4 & 1\,1/4 & 3/4 & 1 \\ s_5 & 0 & 3/4 & 3/4 & 3/4 & 1\,1/4 & 1 \\ s_6 & 0 & 3/4 & 3/4 & 1 & 1 & 1\,3/8 \end{array}$$

som gir

$$A^{-1} = \frac{1}{6}\begin{bmatrix} 6 & 0 & 0 & 0 & 0 & 0 \\ 0 & 14 & 2 & -6 & -6 & 0 \\ 0 & 2 & 14 & -6 & -6 & 0 \\ 0 & -6 & -6 & 16 & 4 & -8 \\ 0 & -6 & -6 & 4 & 16 & -8 \\ 0 & 0 & 0 & -8 & -8 & 16 \end{bmatrix}$$

"Mixed-model"-ligningene blir da:

$$\begin{bmatrix} 25 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 2 & 2 & 2 \\ 0 & 42^{*} & 5 & -15 & -15 & 0 & 3 & 0 & 4 & 0 & 0 & 0 \\ 0 & 5 & 41 & -15 & -15 & 0 & 0 & 2 & 0 & 0 & 0 & 4 \\ 0 & -15 & -15 & 46 & 10 & -20 & 3 & 3 & 0 & 0 & 0 & 0 \\ 0 & -15 & -15 & 10 & 46 & -20 & 0 & 0 & 6 & 0 & 0 & 0 \\ 0 & 0 & 0 & -20 & -20 & 45 & 0 & 0 & 0 & 3 & 0 & 2 \\ 4 & 3 & 0 & 3 & 0 & 0 & 10 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 3 & 0 & 0 & 0 & 5 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 6 & 0 & 0 & 0 & 10 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 5 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \\ 2 & 0 & 4 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 8 \end{bmatrix} \begin{bmatrix} \hat{s}_1 \\ \hat{s}_2 \\ \hat{s}_3 \\ \hat{s}_4 \\ \hat{s}_5 \\ \hat{s}_6 \\ \hat{h}_1 \\ \hat{h}_2 \\ \hat{h}_3 \\ \hat{h}_4 \\ \hat{h}_5 \\ \hat{h}_6 \end{bmatrix} = \begin{bmatrix} 1408 \\ 1040 \\ 1044 \\ 860 \\ 983 \\ 889 \\ 1396 \\ 709 \\ 1607 \\ 828 \\ 321 \\ 1363 \end{bmatrix}$$

$^{*}7 + (\frac{14}{6})(15)) = 42.$

Lösningene blir:

$\hat{s}_1 = -5{,}86436$ $\quad\quad$ $\hat{h}_1 = 139{,}58223$

$\hat{s}_2 = 2{,}89166$ $\quad\quad$ $\hat{h}_2 = 136{,}44602$

$\hat{s}_3 = 5{,}90488$ $\quad\quad$ $\hat{h}_3 = 156{,}33321$

$\hat{s}_4 = 4{,}98672$ $\quad\quad$ $\hat{h}_4 = 164{,}37492$

$\hat{s}_5 = 5{,}35020$ $\quad\quad$ $\hat{h}_5 = 166{,}36436$

$\hat{s}_6 = 5{,}95138$ $\quad\quad$ $\hat{h}_6 = 167{,}40080$

Her kan en se at far-effektene ("sire proofs") ikke summerer til 0 ($G \neq I\sigma_s^2$). Videre kan en sammenligne far-effektene fra eksempel B (dvs. $\widehat{g_i + s_{ij}}$) med far-effektene fra eksempel D.1 ($\hat{s}_j$) for å se om en ville ha selektert feil ved å ignorere slektskapet mellom hanndyrene.

Når det gjelder selve inverteringen av A-matrisen, kan dette bli nokså komplisert i mange tilfeller når matrisene er store. En bör derfor utnytte enkle lösninger om slike finnes. La oss ta et praktisk eksempel.

Vi forutsetter u^* far-sönn grupper, og mellom grupper forutsetter vi at det additive genetiske slektskapet er 0. Dette innebärer at:

$$A = \begin{bmatrix} A_1 & & \\ & A_u & 0 \\ & \cdot & \\ & & \cdot \\ 0 & & A_{u^*} \end{bmatrix}$$

Den u^{te} submatrise har da fölgende sammensetning:

$$A_u = \left[\begin{array}{c|ccccc} 1 & \frac{1}{2} & \frac{1}{2} & \cdot & \cdot & \frac{1}{2} \\ \hline \frac{1}{2} & 1 & \frac{1}{4} & \cdot & \cdot & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{4} & 1 & \cdot & \cdot & \frac{1}{4} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \frac{1}{2} & \frac{1}{4} & \cdot & \cdot & \cdot & 1 \end{array}\right] \quad (v_u + 1) \times (v_u + 1)$$

hvor v_u er antall sönner i u^{te} gruppe.

Dette betyr at $[A_u \; A_u^{-1} = I]$:

$$\begin{bmatrix} 1 & \frac{1}{2}1^{*\prime} \\ \frac{1}{2}1^{*} & \frac{3}{4}I + \frac{1}{4}J \end{bmatrix} \begin{bmatrix} a_1 & a_2 1^{*\prime} \\ a_2 1^{*} & a_3 I + a_4 J \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & I \end{bmatrix}$$

hvor

I er en $v_u \times v_u$ "identity" matrise

J er en $v_u \times v_u$ matrise av enere

1^* er en $v_u \times 1$ vektor av enere

1 er en "scalar"

Löser en så med hensyn til a'ene, blir:

$$a_1 = 1 + \frac{1}{3}v_u$$

$$a_2 = -\frac{2}{3}$$

$$a_3 = \frac{4}{3}$$

$$a_4 = 0$$

hvilket innebärer at:

$$A_u^{-1} = \left[\begin{array}{c|ccccc} 1 + \frac{1}{3}v_u & -\frac{2}{3} & & \cdot & \cdot & -\frac{2}{3} \\ \hline -\frac{2}{3} & \frac{4}{3} & & & & \\ \cdot & & & \cdot & 0 & \\ \cdot & & & & \cdot & \\ \cdot & & & 0 & \frac{4}{3} & \\ -\frac{2}{3} & & & & & \frac{4}{3} \end{array}\right]$$

Denne matrisen er altså bare avhengig av v_u, og videre:

$$A^{-1} = \begin{bmatrix} A_1^{-1} & 0 & . & 0 \\ 0 & A_2^{-1} & . & . \\ . & . & . & . \\ 0 & . & . & A_{u*}^{-1} \end{bmatrix}$$

På et enkelt sett kan altså A-matrisen for dette eksemplet inverteres.

Nå har imidlertid Henderson (1975 c) utviklet en metode som gjör at A^{-1} kan oppnås uten större vanskeligheter. Vi skal kort beskrive metoden:

La A väre matrisen for det additive genetiske slektskapet mellom n individer. De förste t ($< n$) individene utgjör baspopulasjonen (ikke innavlete og ubeslektede individer, foreldre ikke spesifisert). Videre la L väre en nedre triangulär matrise slik at $LL' = A$. Spörsmålet er da hvordan en skal beregne L. Dette avhenger selvsagt av om begge, den ene eller ingen av foreldrene er kjente. En kan summere opp beregningene på fölgende sett:

a. Nummerér individene fra 1 ... n slik at foreldre kommer foran deres avkom.

b. Den övre t^2 submatrisen av L er I.

c. Gitt at p og q ($p < q$) er foreldre til individ i, blir:

$$l_{ij} = (l_{pj} + l_{qj})/2 \text{ for } j = 1, \ldots, p$$

$$= l_{qj}/2 \text{ for } j = p+1, \ldots, q$$

$$= 0 \text{ for } j = q+1, \ldots, i-1, \text{ forutsatt at } q < i-1;$$

$$l_{ii} = (1 + 0{,}5 \sum_{j=1}^{p} l_{pj}\, l_{qj} - \sum_{j=1}^{q} l_{ij}^2)^{0{,}5}$$

d. Om bare den ene av foreldrene er kjent, eksempelvis p_i:

$$l_{ij} = l_{pj}/2 \text{ for } j = 1,\dots, p$$

$$= 0 \text{ for } j = p+1,\dots, i-1, \text{ forutsatt at } p < i-1.$$

$$l_{ii} = (1 - \sum_{j=1}^{n} l_{ij}^2)^{0,5}$$

e. Om ingen av foreldrene er kjente:

$$l_{ij} = 0 \text{ for } \sum_{j=1}^{n},\dots, i-1;$$

$$l_{ii} = 1.$$

Metoden kan best illustreres med et eksempel.

Individ	Foreldre
1	Ukjent
2	Ukjent
3	1 og ukjent
4	1 og 2
5	3 og 4
6	1 og 4
7	5 og 6

La oss tenke oss det datasettet som er vist ovenfor hvor individene 1 og 2 utgjör baspopulasjonen. Hvordan vil da L se ut ? Som tidligere er L den nedre triangulär-matrisen, og vi starter med förste linjen nedenfor baspopulasjonen dvs. 3. linje i vårt eksempel.

3. linje:

$i = 3;\ j = 1;\ p = 1;\ q = \text{ukjent};\ l_{31} = \frac{1}{2}$ } Förste element

$j = 2\ (= p+1);\ l_{32} = 0$ } Andre element

$l_{33} = [1 - (\frac{1}{2}^2 + 0^2)]^{0,5} = 0,866025$ } Tredje element

Disse elementene kan så föres inn i L.

4. linje:

$i = 4;\ j = 1;\ p = 1;\ q = 2;\ l_{41} = \frac{1+0}{2} = \frac{1}{2}$ } Förste element

$j = 2;\ p = 1;\ q = 2;\ l_{42} = \frac{1}{2}$ } Andre element

$j = 3\ (= q+1);\ l_{43} = 0$ } Tredje element

$l_{44} = [1 + 0 - (0,5^2 + 0,5^2)]^{0,5} = 0,5^{0,5}$ } Fjerde element

5. linje:

$i = 5;\ j = 1;\ p = 3;\ q = 4;\ l_{51} = \frac{0,5 + 0,5}{2} = 0,5$ } Förste element

$j = 2;\ p = 3;\ q = 4;\ l_{52} = \frac{0 + 0,5}{2} = 0,25$ } Andre element

$j = 3;\ p = 3;\ q = 4;\ l_{53} = \frac{0,866025 + 0}{2} = 0,433016$ } Tredje element

$j = 4;\ p = 3;\ q = 4;\ l_{54} = \frac{0 + 0,707107}{2} = 0,353553$ } Fjerde element

$$l_{55} = [1 + 0,5\ \left(\frac{(0,5)(0,5) + (0)(0,5) + (0,866025)(0)}{2}\right) -$$

$$(0,5^2 + 0,25^2 + 0,433016^2 + 0,353553^2)]^{0,5} = 0,5^{0,5},$$

som er femte element, osv.

Resultatene er vist i tabellene nedenfor og på neste side hvor også A^{-1} (beregnet på tradisjonelt sett) er tatt med.

L =	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1)	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000
(2)	0,000000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000
(3)	0,500000	0,000000	0,866025	0,000000	0,000000	0,000000	0,000000
(4)	0,500000	0,500000	0,000000	0,707107	0,000000	0,000000	0,000000
(5)	0,500000	0,250000	0,433013	0,353553	0,707107	0,000000	0,000000
(6)	0,750000	0,250000	0,000000	0,353553	0,000000	0,707107	0,000000
(7)	0,625000	0,250000	0,216506	0,353553	0,353553	0,353553	0,637378

$$LL' = A = \begin{array}{cc} & \begin{array}{ccccccc} (1) & (2) & (3) & (4) & (5) & (6) & (7) \end{array} \\ \begin{array}{c} (1) \\ (2) \\ (3) \\ (4) \\ (5) \\ (6) \\ (7) \end{array} & \left[\begin{array}{ccccccc} 1{,}000000 & 0{,}000000 & 0{,}500000 & 0{,}500000 & 0{,}500000 & 0{,}750000 & 0{,}625000 \\ 0{,}000000 & 1{,}000000 & 0{,}000000 & 0{,}500000 & 0{,}250000 & 0{,}250000 & 0{,}250000 \\ 0{,}500000 & 0{,}000000 & 1{,}000000 & 0{,}250000 & 0{,}625000 & 0{,}375000 & 0{,}500000 \\ 0{,}500000 & 0{,}500000 & 0{,}250000 & 1{,}000000 & 0{,}625000 & 0{,}750000 & 0{,}687500 \\ 0{,}500000 & 0{,}250000 & 0{,}625000 & 0{,}625000 & 1{,}125000 & 0{,}562500 & 0{,}843750 \\ 0{,}750000 & 0{,}250000 & 0{,}375000 & 0{,}750000 & 0{,}562500 & 1{,}250000 & 0{,}906250 \\ 0{,}625000 & 0{,}250000 & 0{,}500000 & 0{,}687500 & 0{,}843750 & 0{,}906250 & 1{,}281250 \end{array}\right] \end{array}$$

$$A^{-1} = \begin{array}{cc} & \begin{array}{ccccccc} (1) & (2) & (3) & (4) & (5) & (6) & (7) \end{array} \\ \begin{array}{c} (1) \\ (2) \\ (3) \\ (4) \\ (5) \\ (6) \\ (7) \end{array} & \left[\begin{array}{ccccccc} 2{,}333333 & 0{,}500000 & -0{,}666667 & -0{,}500000 & 0{,}000000 & -1{,}000000 & 0{,}000000 \\ 0{,}500000 & 1{,}500000 & 0{,}000000 & -1{,}000000 & 0{,}000000 & 0{,}000000 & 0{,}000000 \\ -0{,}666667 & 0{,}000000 & 1{,}833333 & 0{,}500000 & -1{,}000000 & -1{,}000000 & 0{,}000000 \\ -0{,}500000 & -1{,}000000 & 0{,}500000 & 3{,}000000 & -1{,}000000 & -1{,}000000 & 0{,}000000 \\ 0{,}000000 & 0{,}000000 & -1{,}000000 & -1{,}000000 & 2{,}615384 & 0{,}615384 & -1{,}230767 \\ -1{,}000000 & 0{,}000000 & 0{,}000000 & -1{,}000000 & 0{,}615384 & 2{,}615384 & -1{,}230767 \\ 0{,}000000 & 0{,}000000 & 0{,}000000 & 0{,}000000 & -1{,}230767 & -1{,}230767 & 2{,}461534 \end{array}\right] \end{array}$$

Den forenklede metoden som Henderson har utviklet bygger på anvendelsen av den inverse verdien av de kvadrerte diagonal-elementene av L, dvs. $d_i = \dfrac{1}{l_{ii}^2}$. For vårt eksempel er:

$$d' = [1 \quad 1 \quad 1{,}333 \quad 2 \quad 2 \quad 2 \quad 2{,}461534].$$

La a^{ij} ($= a^{ji}$) väre ij^{te} element av A^{-1}. Disse elementene kan da beregnes på fölgende måte:

a. $a^{ii} = d_i + 0{,}25 \sum\limits_k d_k$ hvor k refererer seg til avkom etter i^{te} far

Eksempel: $a^{11} = [1 + 0{,}25\ (1{,}333 + 2 + 2)] = 2{,}33$

$a^{44} = [2 + 0{,}25\ (2 + 2)] = 3.$

b. a^{ij} $(i < j) = -0{,}5\ d_j + 0{,}25 \sum_k d_k$ om j er avkom etter i, eller

$0{,}25 \sum_k d_k$ etter paring mellom i og j.

k refererer seg altså til avkom etter ij^{te} paring.

Eksempel: $a^{12} = (0{,}25)(2) = 0{,}50$ (nr. 4 er felles avkom)

$a^{67} = (-0{,}5)(2{,}461534) = -1{,}230767$ (7 er avkom etter 6)

$a^{14} = (-0{,}5)(2) + 0{,}25\ (2) = -0{,}5$ (1 x 4 gir avkom 6)

Som en ser, stemmer disse resultatene med elementene i A^{-1}.

Om en er sikker på at en arbeider med en ikke innavlet populasjon, blir den forenklede metoden enda enklere å anvende. De eneste 3 verdiene som da kan forekomme for d_i er:

1 om ingen av foreldrene til individ i er kjente

$\frac{4}{3}$ om bare den ene av foreldrene til individ i er kjent

2 om begge foreldrene til individ i er kjente

Vi kan ta et eksempel:

Individ	Foreldre
1	Ukjent
2	Ukjent
3	1 og 2
4	1 og 2
5	1 og 2
6	1 og ukjent
7	1 og ukjent
8	2 og ukjent
9	7 og 8

Således: $d' = [1 \quad 1 \quad 2 \quad 2 \quad 2 \quad 4/3 \quad 4/3 \quad 4/3 \quad 2]$, og

$$
A^{-1} = \frac{1}{6}\begin{bmatrix}
19 & 9 & -5 & -6 & -6 & -4 & -4 & 0 & 0 \\
9 & 17 & -6 & -6 & -6 & 0 & 0 & -4 & 0 \\
-6 & -6 & 12 & 0 & 0 & 0 & 0 & 0 & 0 \\
-6 & -6 & 0 & 12 & 0 & 0 & 0 & 0 & 0 \\
-6 & -6 & 0 & 0 & 12 & 0 & 0 & 0 & 0 \\
-4 & 0 & 0 & 0 & 0 & 8 & 0 & 0 & 0 \\
-4 & 0 & 0 & 0 & 0 & 0 & 11 & 3 & -6 \\
0 & -4 & 0 & 0 & 0 & 0 & 3 & 11 & -6 \\
0 & 0 & 0 & 0 & 0 & 0 & -6 & -6 & 12
\end{bmatrix}
$$

Noen illustrasjoner:

$$a^{11} = 1 + 0,25\ [2 + 2 + 2 + \tfrac{4}{3} + \tfrac{4}{3}] = \frac{19}{6}$$

$$a^{12} = 0,25\ [2 + 2 + 2] = \frac{9}{6}$$ [3, 4 og 5 er felles avkom]

$$a^{79} = (-0,5)(2) = -1,0$$ [9 er avkom etter 7].

Invertering av matriser er også diskutert av Henderson (1975 e).

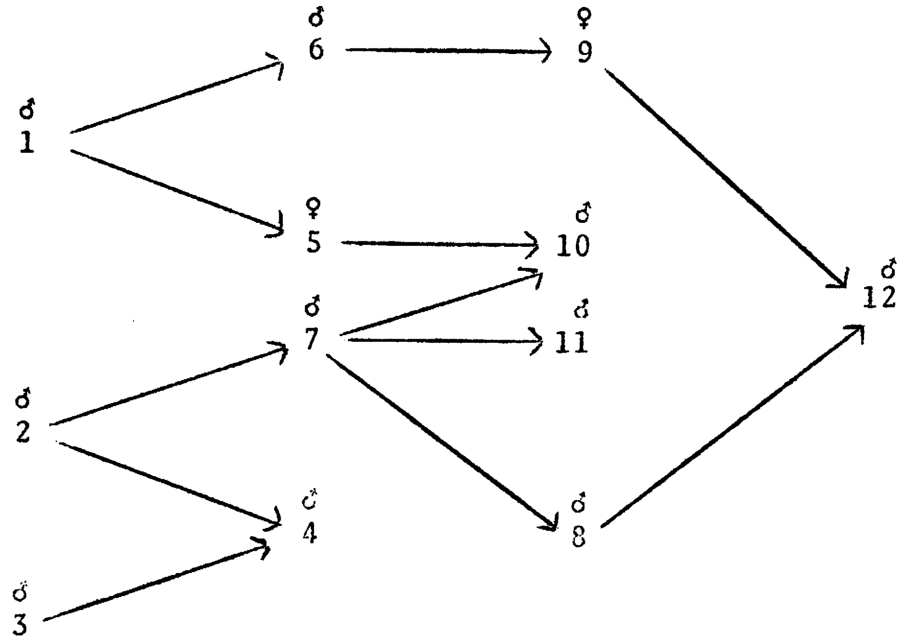
2. Henderson (1975 d)

Vi forutsetter at vi har et materiale som omfatter oksene 1, 2, 3, 4, 6, 7, 8, 10, 11 og 12. Oksene 8 og 12 har ingen avkom med informasjon om mjölkeavkastning. Etter at effekten av besetning-år-sesong er absorbert, tenker vi oss fölgende "least-squares" ligninger for okser med avkom, som har informasjon om mjölkeavkastning:

$$
\begin{array}{cccccccc}
s_1 & s_2 & s_3 & s_4 & s_6 & s_7 & s_{10} & s_{11}
\end{array}
$$

$$
\left[\begin{array}{ccc|ccc|cc}
13,784 & -4,989 & -2,946 & -2,213 & -1,500 & -2,136 & 0,000 & 0,000 \\
-4,989 & 13,261 & -3,476 & -1,382 & -1,429 & -1,985 & 0,000 & 0,000 \\
-2,946 & -3,476 & 6,422 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 \\
\hline
-2,213 & -1,382 & 0,000 & 7,529 & -1,500 & -2,434 & 0,000 & 0,000 \\
-1,500 & -1,429 & 0,000 & -1,500 & 4,429 & 0,000 & 0,000 & 0,000 \\
-2,136 & -1,985 & 0,000 & -2,434 & 0,000 & 12,000 & -1,417 & -4,028 \\
\hline
0,000 & 0,000 & 0,000 & 0,000 & 0,000 & -1,417 & 3,875 & -2,458 \\
0,000 & 0,000 & 0,000 & 0,000 & 0,000 & -4,028 & -2,458 & 6,486
\end{array}\right]
\begin{bmatrix}
\hat{s}_1 \\ \hat{s}_2 \\ \hat{s}_3 \\ \hat{s}_4 \\ \hat{s}_6 \\ \hat{s}_7 \\ \hat{s}_{10} \\ \hat{s}_{11}
\end{bmatrix}
=
\begin{bmatrix}
190 \\ 3916 \\ -3264 \\ -2275 \\ 32 \\ 1826 \\ -1162 \\ 737
\end{bmatrix}
$$

Ingen av hunndyrene (mödre 5 og 9) er testet, og relasjonene mellom de forskjellige individene går frem av diagrammet nedenfor:



Etter som ingen innavl foreligger kan Hendersons forenklede metode brukes for å finne A^{-1}. Dette gir for individene 1 til 12 (bruk kovarians-metoden for beregning av A):

$$A^{-1} = \frac{1}{6}\begin{bmatrix}
10^{*} & 0 & 0 & 0 & -4 & -4 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 10 & 0 & -4 & 0 & 0 & -4 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -4 & 0 & 8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-4 & 0 & 0 & 0 & 11 & 0 & 3 & 0 & 0 & -6 & 0 & 0 \\
-4 & 0 & 0 & 0 & 0 & 10 & 0 & 0 & -4 & 0 & 0 & 0 \\
0 & -4 & 0 & 0 & 3 & 0 & 15 & -4 & 0 & -6 & -4 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -4 & 11 & 3 & 0 & 0 & -6 \\
0 & 0 & 0 & 0 & 0 & -4 & 0 & 3 & 11 & 0 & 0 & -6 \\
0 & 0 & 0 & 0 & -6 & 0 & -6 & 0 & 0 & 12 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -4 & 0 & 0 & 0 & 8 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -6 & -6 & 0 & 0 & 12
\end{bmatrix}$$

$^{*}a^{11} = 1 + 0{,}25\ (\frac{4}{3} + \frac{4}{3}) = \frac{10}{6}$

For den videre bearbeidingen kan det väre rasjonelt å dele opp A^{-1} i submatriser dvs.

$$A^{-1} = \begin{bmatrix} A_{11} & A_{12} \\ A_{12}' & A_{22} \end{bmatrix}^{-1} = \begin{bmatrix} W_{11} & W_{12} \\ W_{12}' & W_{22} \end{bmatrix}$$

A_{11} refererer seg til fedre med avkom som har informasjon om mjölkeavkastning

A_{22} refererer seg til fedre uten registrering av avkastningen på avkommet og mödre uten registrering av avkastningen

Rekker og kolonner må da grupperes om, og BLUP-ligningene kan skrives som:

$$\begin{bmatrix} C_{22} + W_{11}r^* & W_{12}r^* \\ W_{12}'r^* & W_{22}r^* \end{bmatrix} \begin{bmatrix} \hat{s}_1 \\ \hat{s}_2 \end{bmatrix} = \begin{bmatrix} t \\ 0 \end{bmatrix}$$

$\hat{s}_1$ er BLUP for fedre med avkom, og

$\hat{s}_2$ er BLUP for de övrige individene

$r^* = \frac{\sigma_e^2}{\sigma_s^2} = 15$ (bare förste laktasjonsregistreringen er brukt)

På neste side har en vist <u>BLUP-ligningene</u> uten å gruppere om rekker og kolonner [NB! en estimerer halvparten av individenes genotyper]:

En skal for oversiktens skyld vise hvordan en har kommet frem til enkelte koeffisienter:

$\hat{s}_1$ i ligningen for s_1: $13{,}784 + (\frac{10}{6})\ 15 = 38{,}784$

$\hat{s}_6$ i ligningen for s_1: $-1{,}5 - (\frac{4}{6})\ 15 = -11{,}5$

$\hat{s}_5$ i ligningen for s_1: $0 - (\frac{4}{6})\ 15 = -10$ osv.

Lösningene er:

$\hat{s}$ = [16,75 122,11 -130,25 -20,15 -13,24 15,10
111,23 55,61 7,55 20,17 88,61 31,58]

$$\begin{bmatrix} 38,784 & -4,989 & -2,946 & -2,213 & -10,000 & -11,500 & -2,136 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 \\ -4,989 & 38,261 & -3,476 & -11,382 & 0,000 & -1,429 & -11,985 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 \\ -2,946 & -3,476 & 21,422 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 \\ -2,213 & -11,382 & 0,000 & 27,529 & 0,000 & -1,500 & -2,434 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 \\ -10,000 & 0,000 & 0,000 & 0,000 & 27,500 & 0,000 & 7,500 & 0,000 & 0,000 & -15,000 & 0,000 & 0,000 \\ -11,500 & -1,429 & 0,000 & -1,500 & 0,000 & 29,429 & 0,000 & 0,000 & -10,000 & 0,000 & 0,000 & 0,000 \\ -2,136 & -11,985 & 0,000 & -2,434 & 7,500 & 0,000 & 49,500 & -10,000 & 0,000 & -16,417 & -14,028 & 0,000 \\ 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & -10,000 & 27,500 & 7,500 & 0,000 & 0,000 & -15,000 \\ 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & -10,000 & 0,000 & 7,500 & 27,500 & 0,000 & 0,000 & -15,000 \\ 0,000 & 0,000 & 0,000 & 0,000 & -15,000 & 0,000 & -16,417 & 0,000 & 0,000 & 33,875 & -2,458 & 0,000 \\ 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & -14,028 & 0,000 & 0,000 & -2,458 & 26,486 & 0,000 \\ 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & -15,000 & -15,000 & 0,000 & 0,000 & 30,000 \end{bmatrix} \begin{bmatrix} \hat{s}_1 \\ \hat{s}_2 \\ \hat{s}_3 \\ \hat{s}_4 \\ \hat{s}_5 \\ \hat{s}_6 \\ \hat{s}_7 \\ \hat{s}_8 \\ \hat{s}_9 \\ \hat{s}_{10} \\ \hat{s}_{11} \\ \hat{s}_{12} \end{bmatrix} = \begin{bmatrix} 190 \\ 3916 \\ -3264 \\ -2275 \\ 0 \\ 32 \\ 1826 \\ 0 \\ 0 \\ -1162 \\ 737 \\ 0 \end{bmatrix}$$

En kan merke seg her at f.eks.:

$\hat{s}_{12} = \frac{\hat{s}_8 + \hat{s}_9}{2}$, som også er forventet da individ 12 ikke har

noen egen informasjon eller avkom med informasjon om mjölkeavkastning.

Videre er: $1'A^{-1}\hat{s} = 0$, dvs. for vårt eksempel:

$$\frac{1}{6}\,[2 \quad 2 \quad 6 \quad 4 \quad 4 \quad 2 \quad 0 \quad 4 \quad 4 \quad 0 \quad 4 \quad 0]\,\hat{s} = 0,$$

noe som kan väre nyttig å vite for å redusere konvergeringstiden i en eventuell iterasjon.

Om en senere får informasjon om mödrenes (individ 5 og 9) avkastning kan en ifölge Henderson (1975d) inkludere denne på fölgende sett:

- Addér $\frac{n\,h^2\,r^*}{n\,(r - h^2) + (1 - r)}$ til diagonalkoeffisienten for

henholdsvis ku nr. 5 og 9 i de tidligere BLUP-ligningene. Denne formelen har fölgende bakgrunn:

La oss tenke oss fölgende modell for to kyr:

$$y_{ij} = g_i + p_i + e_{ij}\text{, hvor}$$

g_i = genotypen for i^{te} ku for en egenskap

p_i = permanent miljöeffekt for i^{te} ku, osv.

Ku nr. 1 og 2 har henholdsvis n_1 og n_2 registreringer.

Da blir BLUP-ligningene:

$$y_{ij} = \underbrace{g_i + p_i}_{\text{ku-effekt}} + e_{ij}\text{, hvor}$$

$$\begin{bmatrix} n_1 + a^{11}\gamma & a^{12}\gamma & n_1 & 0 \\ a^{12}\gamma & n_2 + a^{22}\gamma & 0 & n_2 \\ n_1 & 0 & n_1 + \alpha & 0 \\ 0 & n_2 & 0 & n_2 + \alpha \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} n_1\bar{y}_1 \\ n_2\bar{y}_2 \\ n_1\bar{y}_1 \\ n_2\bar{y}_2 \end{bmatrix}$$

hvor

$$\alpha = \left[\frac{1 - r}{r - h^2}\right] = \frac{\sigma_e^2}{\sigma_p^2}, \text{ og } \gamma = \frac{\sigma_e^2}{\sigma_g^2}$$

Som tidligere er:

$$A^{-1} = \begin{bmatrix} a^{11} & a^{12} \\ a^{12} & a^{22} \end{bmatrix}$$

Vi absorberer så p_1 og p_2:

$$\begin{bmatrix} n_1 + a^{11}\gamma & a^{12}\gamma \\ a^{12}\gamma & n_2 + a^{22}\gamma \end{bmatrix} - \begin{bmatrix} n_1 & 0 \\ 0 & n_2 \end{bmatrix} \begin{bmatrix} n_1 + \alpha & 0 \\ 0 & n_2 + \alpha \end{bmatrix}^{-1} \begin{bmatrix} n_1 & 0 \\ 0 & n_2 \end{bmatrix}$$

som gir diagonalelementene:

$(n_i + a^{ii}\gamma) - \dfrac{n_i^2}{n_i + \alpha}$ mens ingen endring skjer i "off"-diagonal-elementene.

Diagonal-elementene kan videre vises å väre lik:

$$n_i + a^{ii}\gamma - \frac{n_i^2}{n_i + \left[\frac{1 - r}{r - h^2}\right]} = n_i + a^{ii}\gamma - \frac{n_i^2 (r - h^2)}{n_i (r - h^2) + (1 - r)}$$

$$= a^{ii}\gamma + \frac{n_i[n_i (r - h^2) + (1 - r)] - n_i^2 (r - h^2)}{n_i (r - h^2) + (1 - r)}$$

$$= a^{ii}\gamma + \frac{n_i (1 - r)}{n_i (r - h^2) + (1 - r)} = a^{ii}\gamma + n_i h^2\gamma/[n_i (r - h^2) + (1 - r)],$$

fordi $h^2\gamma = \dfrac{\sigma_g^2}{\sigma_y^2} \dfrac{\sigma_e^2}{\sigma_g^2} = \dfrac{\sigma_e^2}{\sigma_y^2} = \dfrac{\sigma_y^2 - \sigma_g^2 - \sigma_p^2}{\sigma_y^2} = 1 - r.$

Om det er halvparten av kuas genotype vi vil estimere, blir $\gamma = r^*$, og dermed skal vi addere $\frac{n\ h^2\ r^*}{n\ (r - h^2) + (1 - r)}$ til tidligere diagonalkoeffisient.

$\therefore$ Addér $\left[\frac{n\ h^2\ r^*}{n\ (r - h^2) + (1 - r)}\right] \frac{\bar{y}}{2}$ til höyresiden i de tidligere BLUP-ligningene for henholdsvis ku nr. 5 og 9. Denne formelen har fölgende bakgrunn:

For eksemplet med de to kyrne må vi også absorbere höyresiden tilhörende p_1 og p_2, dvs.

$$n_i\bar{y}_i - \left[\frac{n_i}{n_i + \alpha}\right] n_i\bar{y}_i = n_i\bar{y}_i - \frac{n_i^2\ (r - h^2)\bar{y}_i}{n_i\ (r - h^2) + (1 - r)}$$

$$= \frac{n_i\bar{y}_i\ [n_i\ (r - h^2) + (1 - r)] - n_i^2\ (r - h^2)\bar{y}_i}{n_i\ (r - h^2) + (1 - r)}$$

$$= \frac{n_i\ (1 - r)\bar{y}_i}{n_i\ (r - h^2) + (1 - r)} = \frac{n_i\ h^2\gamma\ \bar{y}_i}{n_i\ (r - h^2) + (1 - r)}$$

Om det er halvparten av kuas genotype vi vil estimere, blir $\gamma = r^*$, og vi må multiplisere höyresiden med $\frac{1}{2}$. Fölgelig skulle formelen väre bevist.

EKSEMPEL E. BEREGNING AV FENOTYPISK, GENETISK OG MILJÖMESSIG TREND (SCHAEFFER, 1974)

La oss forutsette fölgende datasett:

Dyr	Tidsperiode 1	2	3	4	Totalt
1	58	68	-	-	126
2 *g_1	110	122	114	-	346
3	123	94	103	107	427
4	-	69	53	-	122
5	- g_2	72	105	62	239
6	-	108	63	-	171
7	-	-	98	112	210
8	-	- g_3	103	121	224
9	-	-	121	129	250
10	-	-	-	91	91
11	-	-	- g_4	89	89

*g_1 = generasjon 1 osv.

Dyrene er ubeslektet og ikke innavlet. En eventuell alderseffekt er korrigert bort, og arvbarheten (h^2) settes lik 0,10. Reproduserbarheten er 0,25 (r).

Hvert år kommer det inn en ny generasjon av dyr.

Modellen for dette materialet kan uttrykkes som:

$$y_{ijk} = t_i + g_j + a_{jk} + e_{ijk} \text{ , hvor}$$

t_i = en effekt felles for alle registreringer i samme tidsperiode ("fixed")

g_j = en genetisk effekt felles for alle individer i generasjon j ("fixed")

a_{jk} = effekten av det enkelte individ innen j^{te} gruppe ("random")

e_{ijk} = tilfeldig feil $(0, \sigma_e^2)$

y_{ijk} = observasjon på k^{te} individ i j^{te} generasjon og i^{te} tidsperiode

Videre er $E(a, e) = 0$.

Om en så deler opp a_{jk}-effekten i en genetisk del (a_g) og en del for permanent miljöeffekt (a_p), får en at:

$$a_{jk} = a_{gjk} + a_{pjk} \; [E(a_{gjk}\, a_{pj'k'}) = 0 \text{ for alle } j, k, j', k']$$

Overförer en dette eksemplet i BLUP-symboler, blir:

$$\underline{b}' = [t_1 \;\dots\; t_4 \quad g_1 \;\dots\; g_4]$$

$$\underline{u}' = [a_{g_{11}} \;\dots\; a_{g_{42}} \; a_{p_{11}} \;\dots\; a_{p_{42}}]$$

$$\text{Var}(\underline{u}) = G\sigma_e^2 = \begin{bmatrix} I\,\dfrac{\sigma_g^2}{\sigma_e^2} & 0 \\ 0 & I\,\dfrac{\sigma_p^2}{\sigma_e^2} \end{bmatrix} \sigma_e^2$$

Videre:

$$\sigma_y^2 = \sigma_a^2 + \sigma_e^2 \;;\; r = \frac{\sigma_a^2}{\sigma_y^2} \;;\; \sigma_a^2 = \frac{r}{(1-r)}\sigma_e^2$$

$$\sigma_a^2 = \sigma_g^2 + \sigma_p^2 \;;\; h^2 = \frac{\sigma_g^2}{\sigma_y^2} \;;\; \sigma_g^2 = \frac{h^2}{(1-r)}\sigma_e^2$$

$$\sigma_p^2 = \sigma_a^2 - \sigma_g^2 = \left[\frac{r}{(1-r)} - \frac{h^2}{(1-r)}\right]\sigma_e^2 = \left[\frac{r-h^2}{1-r}\right]\sigma_e^2$$

Derfor:

$$G\sigma_e^2 = \begin{bmatrix} I(h^2/(1-r)) & 0 \\ 0 & I((r-h^2)/(1-r)) \end{bmatrix} \sigma_e^2$$

$$\text{og } G^{-1}\frac{1}{\sigma_e^2} = \begin{bmatrix} I((1-r)/h^2 & 0 \\ 0 & I((1-r)/(r-h^2)) \end{bmatrix} \frac{1}{\sigma_e^2}$$

$$= \begin{bmatrix} I(7,5) & 0 \\ 0 & I(5,0) \end{bmatrix}$$

En kan så som vist på foregående side addere 7,5 og 5,0 til diagonalene for henholdsvis a_g og a_p ligningene. Ligningene er ikke "confounded" lenger da. For å få en unik lösning må vi imidlertid innföre en "constraint", eks. $\hat{g}_1 = 0$. Nedenfor er "least-squares" ligningene (ikke BLUP-ligningene) vist (NB! ligningene for a_{gjk} er identiske med ligningene for a_{pjk}):

$$\left[\begin{array}{ccccccccccccccccccccccccccccc}
3&0&0&0&0&0&0&1&1&1&0&0&0&0&0&0&0&0&1&1&1&0&0&0&0&0&0&0&0\\
&6&0&0&3&0&0&1&1&1&1&1&1&0&0&0&0&0&1&1&1&1&1&1&0&0&0&0&0\\
&&8&0&3&3&0&0&1&1&1&1&1&1&1&1&0&0&0&1&1&1&1&1&1&1&1&0&0\\
&&&7&1&3&2&0&0&1&0&1&0&1&1&1&1&1&0&0&1&0&1&0&1&1&1&1&1\\
&&&&7&0&0&0&0&0&2&3&2&0&0&0&0&0&0&0&0&2&3&2&0&0&0&0&0\\
&&&&&6&0&0&0&0&0&0&0&2&2&2&0&0&0&0&0&0&0&0&2&2&2&0&0\\
&&&&&&2&0&0&0&0&0&0&0&0&0&1&1&0&0&0&0&0&0&0&0&0&1&1\\
&&&&&&&2&0&0&0&0&0&0&0&0&0&0&2&0&0&0&0&0&0&0&0&0&0\\
&&&&&&&&3&0&0&0&0&0&0&0&0&0&0&3&0&0&0&0&0&0&0&0&0\\
&&&&&&&&&4&0&0&0&0&0&0&0&0&0&0&4&0&0&0&0&0&0&0&0\\
&&&&&&&&&&2&0&0&0&0&0&0&0&0&0&0&2&0&0&0&0&0&0&0\\
&&&&&&&&&&&3&0&0&0&0&0&0&0&0&0&0&3&0&0&0&0&0&0\\
&&&&&&&&&&&&2&0&0&0&0&0&0&0&0&0&0&2&0&0&0&0&0\\
&&&&&&&&&&&&&2&0&0&0&0&0&0&0&0&0&0&2&0&0&0&0\\
\multicolumn{14}{l}{\text{symmetrisk}}&2&0&0&0&0&0&0&0&0&0&0&2&0&0&0\\
&&&&&&&&&&&&&&&2&0&0&0&0&0&0&0&0&0&0&2&0&0\\
&&&&&&&&&&&&&&&&1&0&0&0&0&0&0&0&0&0&0&1&0\\
&&&&&&&&&&&&&&&&&1&0&0&0&0&0&0&0&0&0&0&1\\
&&&&&&&&&&&&&&&&&&2&0&0&0&0&0&0&0&0&0&0\\
&&&&&&&&&&&&&&&&&&&3&0&0&0&0&0&0&0&0&0\\
&&&&&&&&&&&&&&&&&&&&4&0&0&0&0&0&0&0&0\\
&&&&&&&&&&&&&&&&&&&&&2&0&0&0&0&0&0&0\\
&&&&&&&&&&&&&&&&&&&&&&3&0&0&0&0&0&0\\
&&&&&&&&&&&&&&&&&&&&&&&2&0&0&0&0&0\\
&&&&&&&&&&&&&&&&&&&&&&&&2&0&0&0&0\\
&&&&&&&&&&&&&&&&&&&&&&&&&2&0&0&0\\
&&&&&&&&&&&&&&&&&&&&&&&&&&2&0&0\\
&&&&&&&&&&&&&&&&&&&&&&&&&&&1&0\\
&&&&&&&&&&&&&&&&&&&&&&&&&&&&1
\end{array}\right]
\begin{bmatrix} t_1\\ t_2\\ t_3\\ t_4\\ g_2\\ g_3\\ g_4\\ a_{g11}\\ a_{g12}\\ a_{g13}\\ a_{g21}\\ a_{g22}\\ a_{g23}\\ a_{g31}\\ a_{g32}\\ a_{g33}\\ a_{g41}\\ a_{g42}\\ a_{p11}\\ a_{p12}\\ a_{p13}\\ a_{p21}\\ a_{p22}\\ a_{p23}\\ a_{p31}\\ a_{p32}\\ a_{p33}\\ a_{p41}\\ a_{p42} \end{bmatrix}
=
\begin{bmatrix} 291\\ 533\\ 760\\ 711\\ 532\\ 684\\ 180\\ 126\\ 346\\ 427\\ 122\\ 239\\ 171\\ 210\\ 224\\ 250\\ 91\\ 89\\ 126\\ 346\\ 427\\ 122\\ 239\\ 171\\ 210\\ 224\\ 250\\ 91\\ 89 \end{bmatrix}$$

samt BLUE og BLUP på de ulike effektene:

	Lösning	Diagonal (D)*
Generasjon		
1	0,0	0,0
2	-21,9453	0,5477
3	16,9205	0,6715
4	-9,0448	1,2234
Tidsperiode		
1	97,0000	0,4444
2	99,8060	0,3591
3	95,1141	0,4080
4	99,0448	0,5568

Individ	Genetisk	(D)*	Permanent miljöeffekt	(D)*
1	-5,6645	0,1186	-8,4967	0,1668
2	3,6053	0,1159	5,4080	0,1608
3	2,0592	0,1155	3,0887	0,1598
4	-2,3224	0,1187	-3,4835	0,1671
5	0,7247	0,1175	1,0871	0,1644
6	1,5977	0,1187	2,3964	0,1671
7	-1,4400	0,1191	-2,1600	0,1680
8	-0,3200	0,1191	-0,4800	0,1680
9	1,7600	0,1191	2,6400	0,1680
10	0,1000	0,1267	0,1500	0,1850
11	-0,1000	0,1267	-0,1500	0,1850

*Diagonalene i den inverse matrisen

Legg merke til at:

- ° $\hat{a}_{gjk} = \sigma_g^2/\sigma_p^2 \ \hat{a}_{pjk} = 2/3 \ \hat{a}_{pjk}$
- °° Sum av $\hat{a}_{gjk}$ = Sum av $\hat{a}_{pjk}$ = 0 for hver j

∘•∘ The "estimated transmitting ability" (ETA) for hvert individ er lik: $(\hat{a}_{gjk} + \hat{g}_j)$

∘∘∘• The "real producing ability" (RPA) for hvert individ er lik: $\hat{a}_{jk} = \hat{g}_j + \hat{a}_{gjk} + \hat{a}_{pjk}$

∘∘∘∘• Miljötrend og genetisk trend kan estimeres. Således:

Individ	ETA	RPA
1	-5,6645	-14,1612
2	3,6053	9,0133
3	2,0592	5,1479
4	-24,2677	-27,7512
5	-21,2206	-20,1335
6	-20,3476	-17,9512
7	15,4805	13,3205
8	16,6005	16,1205
9	18,6805	21,3205
10	-8,9448	-8,7948
11	-9,1448	-9,2948

Eksempelvis for dyr nr. 4:

ETA = -21,9453 + (-2,3224) = -24,2677

RPA = -21,9453 + (-2,3224) + (-3,4835) = -27,7512

Om vi går ut fra t_1 som basis blir:

∘ Fenotypisk trend lik trenden i middeltallene for periodene

∘• Temporär miljötrend blir trenden i $\hat{t}_i$

∘∘∘ Genetisk trend kan beregnes ut fra:

$$\frac{\sum_j \sum_k (\hat{g}_j + \hat{a}_{gjk})_i}{n_{i\cdot\cdot}} \quad \text{for hver periode}$$

.... Permanent miljötrend lik $\frac{\sum_j \sum_k (\hat{a}_{pjk})_i}{n_{i\cdot\cdot}}$ for hver periode

Periode	Fenotypisk	Temporär	Genetisk	Permanent
1	0,0	0,0	0,0	0,0
2	-8,1667	+2,8060	-10,9727**	0,0
3	-2,0000	-1,8859	-1,1762	+1,0621*
4	+4,5714	+2,0048	+1,9301	+0,5965

dvs. Fenotypisk = temporär + genetisk + permanent.

For eksempel betyr dette:

$$* \quad \frac{5{,}4080 + 3{,}0887 + \ldots + (-0{,}4800) + 2{,}6400}{8} = \underline{1{,}0621}$$

$$** \quad \frac{0{,}0 + (-5{,}6645) + 0{,}0 + 3{,}6053 + 0{,}0 + 2{,}0592 + (-21{,}9453) + (-2{,}3224) + (-21{,}9453) + 0{,}7247 + (-21{,}9453) + 1{,}5977}{6}$$

$$= \underline{-10{,}9727}$$

EKSEMPEL F. BEREGNING AV FAR- OG MOREFFEKT NÅR EN SAMTIDIG TAR HENSYN TIL MATERNAL EFFEKT (SCHAEFFER, 1974)

En kan tenke seg fölgende sett av data:

Fedre	Mödre	Tidsperioder 1	2	3	4
A	1	10		-2	
	2		2		-6
	3	16			-5
	4		6		
B	1		-4		
	2	-2			
	3		-4		
	4	4		-13	
C	4				-4
	5			7	
	6			-12	
D	2			14	
	3			-4	
	5				6
	6				-6

Modellen forutsettes å väre:

$$y_{ijk} = s_i + d_j + m_j + t_k + e_{ijk}$$

hvor

s_i = effekt av i^{te} far $(0\ ,\ \frac{1}{4}\ \sigma_g^2)$

d_j = effekt av j^{te} mor $(0\ ,\ \frac{1}{4}\ \sigma_g^2)$

m_j = maternal effekt $(0\ ,\ \sigma_m^2)$

t_k = effekt av k^{te} tidsperiode ("fixed")

e_{ijk} = tilfeldig fel $(0\ ,\ \sigma_e^2)$

y_{ijk} = observasjon på avkom etter i^{te} far og j^{te} mor i k^{te} tidsperiode

Matrisen (A) for det additive genetiske slektskapet mellom mödre og fedre forutsettes å väre:

	A	B	C	D	1	2	3	4	5	6
A	1	0	$\frac{1}{2}$	$\frac{1}{2}$	0	0	0	0	0	0
B	0	1	0	0	0	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$
C	$\frac{1}{2}$	0	1	$\frac{1}{4}$	$\frac{1}{2}$	0	0	0	0	0
D	$\frac{1}{2}$	0	$\frac{1}{4}$	1	0	$\frac{1}{2}$	0	0	0	0
1	0	0	$\frac{1}{2}$	0	1	0	0	0	0	0
2	0	0	0	$\frac{1}{2}$	0	1	0	0	0	0
3	0	0	0	0	0	0	1	0	$\frac{1}{2}$	0
4	0	0	0	0	0	0	0	1	0	$\frac{1}{2}$
5	0	$\frac{1}{2}$	0	0	0	0	$\frac{1}{2}$	0	1	$\frac{1}{4}$
6	0	$\frac{1}{2}$	0	0	0	0	0	$\frac{1}{2}$	$\frac{1}{4}$	1

Overfört i BLUP-symboler får en:

$$\underline{b}' = (t_1\ t_2\ t_3\ t_4)$$

$$\underline{u}' = (S_A\ S_B\ S_C\ S_D\ \ d_1\ d_2\ d_3\ d_4\ d_5\ d_6\ \ m_1\ m_2\ m_3\ m_4\ m_5\ m_6),$$

og

$$G\sigma_e^2 = \text{Var}\ (\underline{u}) = \begin{bmatrix} A\frac{\sigma_s^2}{\sigma_e^2} & 0 \\ 0 & I\frac{\sigma_m^2}{\sigma_e^2} \end{bmatrix} \sigma_e^2$$

$R = \text{Var}\ (\underline{e}) = I\sigma_e^2$. Den inverse matrisen av G er:

$$G^{-1}\frac{1}{\sigma_e^2} = \begin{bmatrix} A^{-1}\frac{\sigma_e^2}{\sigma_s^2} & 0 \\ 0 & I\frac{\sigma_e^2}{\sigma_m^2} \end{bmatrix} \frac{1}{\sigma_e^2}$$

hvor,

$\sigma_e^2/\sigma_s^2 = \sigma_e^2/\sigma_d^2 = 4$, $\sigma_e^2/\sigma_m^2 = 10$, og A^{-1} er:

	A	B	C	D	1	2	3	4	5	6
A	2	0	-1	-1	0,5	0,5	0	0	0	0
B	0	2	0	0	0	0	0,5	0,5	-1	-1
C	-1	0	2	0	-1	0	0	0	0	0
D	-1	0	0	2	0	-1	0	0	0	0
1	0,5	0	-1	0	1,5	0	0	0	0	0
2	0,5	0	0	-1	0	1,5	0	0	0	0
3	0	0,5	0	0	0	0	1,5	0	-1	0
4	0	0,5	0	0	0	0	0	1,5	0	-1
5	0	-1	0	0	0	0	-1	0	2	0
6	0	-1	0	0	0	0	0	-1	0	2

En får da fölgende BLUP-ligninger:

$$\left[\begin{array}{cccccccccccccccccccc}
4 & 0 & 0 & 0 & 2 & 2 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\
 & 4 & 0 & 0 & 2 & 2 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\
 & & 6 & 0 & 1 & 1 & 2 & 2 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
 & & & 5 & 2 & 0 & 1 & 2 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 \\
 & & & & 15^* & 0 & -4 & -4 & 4 & 4 & 2 & 1 & 0 & 0 & 2 & 2 & 2 & 1 & 0 & 0 \\
 & & & & & 13 & 0 & 0 & 1 & 1 & 3 & 4 & -4 & -4 & 1 & 1 & 1 & 2 & 0 & 0 \\
 & & & & & & 11 & 0 & -4 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\
 & & & & & & & 12 & 0 & -3 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\
 & & & & & & & & 9 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 \\
 & & & & & & & & & 10 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 \\
 & & & & & & & & & & 10 & 0 & -4 & 0 & 0 & 0 & 4 & 0 & 0 & 0 \\
 & & & & & & & & & & & 10 & 0 & -4 & 0 & 0 & 0 & 4 & 0 & 0 \\
 & & & & & & & & & & & & 10 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \\
 & & & & & & & & & & & & & 10 & 0 & 0 & 0 & 0 & 0 & 2 \\
 & & & & & & & & & & & & & & 13 & 0 & 0 & 0 & 0 & 0 \\
 & & & & & & & & & & & & & & & 14 & 0 & 0 & 0 & 0 \\
 & & & & & & & & & & & & & & & & 14 & 0 & 0 & 0 \\
 & & & & & & & & & & & & & & & & & 14 & 0 & 0 \\
 & & & & & & & & & & & & & & & & & & 12 & 0 \\
 & & & & & & & & & & & & & & & & & & & 12
\end{array}\right]
\left[\begin{array}{c} t_1 \\ t_2 \\ t_3 \\ t_4 \\ S_A \\ S_B \\ S_C \\ S_D \\ d_1 \\ d_2 \\ d_3 \\ d_4 \\ d_5 \\ d_6 \\ m_1 \\ m_2 \\ m_3 \\ m_4 \\ m_5 \\ m_6 \end{array}\right]
=
\left[\begin{array}{r} 28 \\ 0 \\ -10 \\ -15 \\ 21 \\ -19 \\ -9 \\ 10 \\ 4 \\ 8 \\ 3 \\ -7 \\ 13 \\ -18 \\ 4 \\ 8 \\ 3 \\ -7 \\ 13 \\ -18 \end{array}\right]$$

$^*7 + (2)(4) = 15$

"Least-squares"-estimatene er gitt nedenfor:

	Estimat	De inverse diagonalelement
Tidsperiode		
1	7,3981	0,4609
2	0,3981	0,4609
3	-2,7338	0,3876
4	-5,1792	0,4457
Fedre		
1	2,1487	0,1757
2	-3,1642	0,1685
3	1,1261	0,1818
4	3,3262	0,1750
Mödre		
1	-0,1784	0,1777
2	1,3119	0,1572
3	0,8064	0,1689
4	-0,9245	0,1708
5	0,4434	0,1834
6	-2,9453	0,1804
Maternale effekter, mödre		
1	-0,1277	0,0880
2	-0,1136	0,0868
3	-0,3263	0,0860
4	-0,0094	0,0860
5	1,2978	0,0892
6	-0,7207	0,0891

Fölgelig kan den genetiske trenden beregnes på fölgende måte:

$$\frac{\sum_i (\hat{s}_i)_k + \sum_j (\hat{d}_j)_k}{n_{..k}} \quad \text{fra år k til k'}$$

Eksempelvis: Den genetiske endringen fra tidsperiode 2 til 1 er:

Periode 2: $\dfrac{2,1487+1,3119+2,1487-0,9245-3,1642-0,1784-3,1642+0,8064}{8} = \underline{-0,127}$

Periode 1: $\dfrac{2,1487-0,1784+2,1487+0,8064-3,1642+1,3119-3,1642-0,9245}{8} = \underline{-0,127}$

Endring: $\underline{0}$

Denne metoden å beregne genetisk endring på innebärer ingen forutsetning om lineäritet. Regresjonsmetoden (f.eks. Smith, , 1962) forutsetter som kjent lineäritet. Selvsagt kan en undersöke om lineäritet foreligger ved å teste forskjellene mellom påfölgende år. Om disse ikke er signifikant forskjellige, kan en beregne en regresjonskoeffisient basert på gjennomsnittene for periodene veid med de inverse verdiene av samplingsvariansene.

Ellers kan en merke seg at verken fareffektene eller moreffektene summerer til null ettersom G ikke er en diagonalmatrise. De maternale effektene summerer imidlertid til null (G er diagonal).

EKSEMPEL G. BRUK AV BLUP-METODEN NÅR EN ARBEIDER MED FLERE EGENSKAPER (SCHAEFFER, 1974)

I dette eksemplet skal vi arbeide med data fra kjöttfé. Tilsammen er det tre fedre. På det hunnlige avkommet er vekt ved avvenning og vekt ved 1 års alder registrert. Alle avkom har informasjon om vekt ved avvenning.

Fedre	Avkom	Vekt ved avvenning Karakter 1	Vekt ved 1 års alder Karakter 2
1	1	406	---
	2	505	747
	3	526	911
	4	417	---
2	1	496	1008
	2	311	---
	3	481	828
	4	455	---
	5	552	752
	6	520	749
3	1	385	---
	2	444	642
	3	469	618
	4	603	833
	5	545	748
	6	470	567
	7	508	665
	8	543	759

La modellen väre:

$$y_{ijk} = \mu_i + s_{ij} + e_{ijk}$$

hvor

i betegner karakter 1 og 2.

$\begin{bmatrix} y_{1jk} \\ y_{2jk} \end{bmatrix}$ = observasjoner på de to egenskapene for k^{te} avkom etter j^{te} far

$\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$ = konstanter (gjennomsnitt) for karakter 1 og 2

$\begin{bmatrix} s_{1j} \\ s_{2j} \end{bmatrix}$ = j^{te} fareffekt ("random") for karakter 1 og 2

Nedenfor har en vist en del av "least-squares"-ligningene:

$$\underset{\underset{\sim}{y}}{\begin{bmatrix} y_{111} \\ y_{112} \\ y_{212} \\ y_{113} \\ y_{213} \\ y_{114} \\ y_{121} \\ y_{221} \\ y_{122} \\ \vdots \end{bmatrix}} = \underset{X}{\begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ & \vdots \end{bmatrix}} \underset{\underset{\sim}{b}}{\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}} + \underset{Z}{\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ & & & \vdots & & \end{bmatrix}} \underset{\underset{\sim}{u}}{\begin{bmatrix} s_{11} \\ s_{21} \\ s_{12} \\ s_{22} \\ s_{13} \\ s_{23} \end{bmatrix}} + \underset{\underset{\sim}{e}}{\begin{bmatrix} e_{111} \\ e_{112} \\ e_{212} \\ e_{113} \\ e_{213} \\ e_{114} \\ e_{121} \\ e_{221} \\ e_{122} \\ \vdots \end{bmatrix}}$$

avkom

far

egenskap

Videre:

$E(s_{1j}) = E(s_{2j}) = 0$. For hver j har vi at:

$$\text{Var.Kov.} \begin{bmatrix} s_{1j} \\ s_{2j} \end{bmatrix} = \begin{bmatrix} 500 & 320 \\ 320 & 800 \end{bmatrix}$$

Er $j \neq j'$, settes kovariansene lik 0.

$$\begin{bmatrix} e_{1jk} \\ e_{2jk} \end{bmatrix}$$ = en tilfeldig effekt for egenskap 1 og 2

$$E \begin{bmatrix} e_{1jk} \\ e_{2jk} \end{bmatrix} = 0 \text{ og}$$

$$\text{Var.Kov.} \begin{bmatrix} e_{1jk} \\ e_{2jk} \end{bmatrix} = \begin{bmatrix} 3500 & 3000 \\ 3000 & 5200 \end{bmatrix}$$

Er $jk \neq j'k'$, settes kovariansene lik 0. Fölgelig,
$R = \text{Var.}(\underline{e}) \neq I\sigma_e^2$ og $G = \text{Var.}(\underline{s}) \neq I\sigma_s^2$.

$$G = \begin{bmatrix} 500 & 320 & 0 & 0 & 0 & 0 \\ 320 & 800 & 0 & 0 & 0 & 0 \\ 0 & 0 & 500 & 320 & 0 & 0 \\ 0 & 0 & 320 & 800 & 0 & 0 \\ 0 & 0 & 0 & 0 & 500 & 320 \\ 0 & 0 & 0 & 0 & 320 & 800 \end{bmatrix}$$

Dimensjonen på G avhenger nå både av antall fedre og antall egenskaper.

Når det gjelder R-matrisen, består denne av blokker på dimensjon 1 eller 2 avhengig av om avkommet har informasjon om bare den ene egenskapen eller begge egenskapene. For far 1 blir submatrisen:

$$R_{\text{Far 1}} = \begin{bmatrix} 3500 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3500 & 3000 & 0 & 0 & 0 \\ 0 & 3000 & 5200 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3500 & 3000 & 0 \\ 0 & 0 & 0 & 3000 & 5200 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3500 \end{bmatrix}$$

For å sette opp BLUP-ligningene er det nödvendig å kjenne til:

$$P_1 = \begin{bmatrix} (3500)^{-1} & 0 \\ 0 & 0 \end{bmatrix}$$

$$P_{12} = \begin{bmatrix} 3500 & 3000 \\ 3000 & 5200 \end{bmatrix}^{-1} = \frac{1}{92000} \begin{bmatrix} 52 & -30 \\ -30 & 35 \end{bmatrix}$$

og

$$G^{-1} = \begin{bmatrix} 500 & 320 \\ 320 & 800 \end{bmatrix}^{-1} = \begin{bmatrix} 0{,}0026882 & -0{,}0010753 \\ -0{,}0010753 & 0{,}0016801 \end{bmatrix}$$

BLUP-ligningene er vist nedenfor:

$$\begin{array}{c} \\ B_0 \\ B_1^* \longrightarrow \\ \\ \\ \\ \\ \\ \end{array}
\begin{bmatrix}
0{,}0088 & -0{,}0042 & 0{,}0017 & -0{,}0007 & 0{,}0028 & -0{,}0013 & 0{,}0042 & -0{,}0023 \\
 & 0{,}0049 & -0{,}0007 & 0{,}0008 & -0{,}0013 & 0{,}0015 & -0{,}0023 & 0{,}0027 \\
 & & 0{,}0044 & -0{,}0017 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 \\
 & & & 0{,}0024 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 \\
 & & & & 0{,}0055 & -0{,}0024 & 0{,}0 & 0{,}0 \\
 & & & & & 0{,}0032 & 0{,}0 & 0{,}0 \\
 & & & & & & 0{,}0069 & -0{,}0034 \\
 & & & & & & & 0{,}0043
\end{bmatrix}
\begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \\ \hat{s}_{11} \\ \hat{s}_{21} \\ \hat{s}_{12} \\ \hat{s}_{22} \\ \hat{s}_{13} \\ \hat{s}_{23} \end{bmatrix}
=
\begin{array}{c} C_1 \quad C_0 \\ \begin{bmatrix} 1{,}1250 \\ 1{,}5661 \\ 0{,}2772 \\ 0{,}2946 \\ 0{,}2888 \\ 0{,}6014 \\ 0{,}5590 \\ 0{,}6702 \end{bmatrix} \end{array}$$

En skal for oversiktens skyld vise hvordan en har kommet frem til de enkelte ledd i disse ligningene:

$$B_j = m_{1j}P_1 + m_{12j}P_{12} \text{ hvor } (j \geq 1)$$

m_{1j} = antall avkom etter j^{te} far med bare 1 registrering

m_{12j} = antall avkom etter j^{te} far med 2 registreringer (dvs. karakter 1 og 2)

For $j = 1$ blir:

$$B_1 = m_{11}P_1 + m_{121}P_{12} =$$

$$(2) \begin{bmatrix} (3500)^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \frac{2}{92000} \begin{bmatrix} 52 & -30 \\ -30 & 35 \end{bmatrix}$$

Ved å addere til G^{-1} får en:

$$\begin{bmatrix} 0{,}0044 & -0{,}0017 \\ -0{,}0017 & 0{,}0024 \end{bmatrix} = B_1^* \quad \text{(se BLUP-ligningene)}$$

Videre,

$$B_0 = \sum_j B_j \ (\text{ikke } \sum_j B_j^*) \text{ , og}$$

$$C_j = P_1 \begin{bmatrix} \sum_{k=1}^{m_1} y_{1jk} \\ 0 \end{bmatrix} + P_{12} \begin{bmatrix} \sum_{k=1}^{m_{12}} y_{1jk} \\ y_{2jk} \end{bmatrix} \quad (j \geq 1) \text{ , hvor}$$

$\sum_{k=1}^{m_1} y_{1jk}$ = summen av registreringene for avkom etter j^{te} far med informasjon bare på <u>karakter 1</u>

$\sum_{k=1}^{m_{12}} y_{ijk}$ = summen av registreringene på i^{te} karakter for avkom etter j^{te} far med informasjon om <u>begge</u> karakterene

Eksempelvis når j = 1:

$$C_1 = \frac{1}{3500} \begin{bmatrix} 406 + 417 \\ 0 \end{bmatrix} + \frac{1}{92000} \begin{bmatrix} 52 & -30 \\ -30 & 35 \end{bmatrix} \begin{bmatrix} \overbrace{505 + 526}^{1031} \\ \underbrace{747 + 911}_{1658} \end{bmatrix} =$$

$$\begin{bmatrix} 0{,}2351 \\ 0 \end{bmatrix} + \begin{bmatrix} 0{,}0421 \\ 0{,}2946 \end{bmatrix} = \begin{bmatrix} 0{,}2772 \\ 0{,}2946 \end{bmatrix}$$

Fölgelig $C_0 = \sum_j C_j$, dvs. $\begin{bmatrix} 0{,}2772 + 0{,}2888 + 0{,}5590 = 1{,}1250 \\ 0{,}2946 + 0{,}6014 + 0{,}6702 = 1{,}5661 \end{bmatrix}$

Ved å löse BLUP-ligningene med hensyn på μ og s får vi:

	Lösning	Diagonalelementene fra den inverse matrisen
$\hat{\mu}_1$	477,4602	367,9868
$\hat{\mu}_2$	740,2704	639,0941
$\hat{s}_{11}$	-7,0518	365,1387
$\hat{s}_{21}$	12,5038	621,7824
$\hat{s}_{12}$	-6,7544	345,8665
$\hat{s}_{22}$	25,4733	574,9301
$\hat{s}_{13}$	13,8062	336,2604
$\hat{s}_{23}$	-37,9771	554,1180

Legg merke til at:

$$\hat{s}_{11} + \hat{s}_{12} + \hat{s}_{13} = 0$$
$$\hat{s}_{21} + \hat{s}_{22} + \hat{s}_{23} = 0$$

Den estimerte ökonomiske avlsverdien blir da:

$\sum_i \hat{v}_i \, (\hat{s}_{ij})$, hvor $\hat{v}_i$ er den beregnete ökonomiske

vekten for i^{te} karakter.

EKSEMPEL H. BRUK AV BLUP-METODEN PÅ ET MER GENERELT PROBLEM (SCHAEFFER, 1974)

Fölgende datasett forutsettes:

	Besetning			Besetning		
Far	1	2	Total	1	2	Total
1	3	5	8	174	784	958
2	2	0	2	58	0	58
3	1	4	5	12	448	460
	6	9	15	244	1232	1476
	n_{ij}			$y_{ij.}$		

Model: $\underline{y} = X\underline{b} + Z\underline{u} + \underline{e}$

$$\begin{bmatrix} 62 \\ 58 \\ 54 \\ 160 \\ 156 \\ 152 \\ 160 \\ 156 \\ 25 \\ 33 \\ 12 \\ 109 \\ 117 \\ 113 \\ 109 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix} + \underline{e}$$

$$E \begin{bmatrix} \underline{y} \\ \underline{u} \\ \underline{e} \end{bmatrix} = \begin{bmatrix} X\underline{b} + ZP\underline{b} + T\underline{b} \\ P\underline{b} \\ T\underline{b} \end{bmatrix}$$

hvor,

$$P = \begin{bmatrix} -0,50 & 0,50 \\ -0,25 & 0,25 \\ -0,10 & 0,10 \end{bmatrix}$$

$$T' = \begin{bmatrix} -,05 & -,05 & -,05 & -,03 & -,03 & -,03 & -,03 & -,03 & -,04 & -,04 & -,05 & -,06 & -,06 & -,06 & -,06 \\ ,05 & ,05 & ,05 & ,03 & ,03 & ,03 & ,03 & ,03 & ,04 & ,04 & ,05 & ,06 & ,06 & ,06 & ,06 \end{bmatrix}$$

$$\text{Var.-Kov.} \begin{bmatrix} \underline{y} \\ \underline{u} \\ \underline{e} \end{bmatrix} = \begin{bmatrix} V & C_1 & C_3 \\ C_1' & G & S \\ C_3' & S' & R \end{bmatrix}$$ (sml. IV 9) og

$$G = \begin{bmatrix} 0,6 & 0,3 & 0,0 \\ 0,3 & 0,6 & 0,0 \\ 0,0 & 0,0 & 0,9 \end{bmatrix} .$$

Videre:

$$S = \begin{bmatrix} 0,1 & 0,1 & 0,1 & 0,1 & 0,1 & 0,1 & 0,1 & 0,1 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 \\ 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,1 & 0,1 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 \\ 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,2 & 0,2 & 0,2 & 0,2 & 0,2 \end{bmatrix}$$

$$R = \begin{bmatrix} 1,0 & 0,3 & 0,3 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 \\ 0,3 & 1,0 & 0,3 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 \\ 0,3 & 0,3 & 1,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 \\ 0,0 & 0,0 & 0,0 & 1,0 & 0,3 & 0,3 & 0,3 & 0,3 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 \\ 0,0 & 0,0 & 0,0 & 0,3 & 1,0 & 0,3 & 0,3 & 0,3 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 \\ 0,0 & 0,0 & 0,0 & 0,3 & 0,3 & 1,0 & 0,3 & 0,3 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 \\ 0,0 & 0,0 & 0,0 & 0,3 & 0,3 & 0,3 & 1,0 & 0,3 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 \\ 0,0 & 0,0 & 0,0 & 0,3 & 0,3 & 0,3 & 0,3 & 1,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 \\ 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 1,0 & 0,3 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 \\ 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,3 & 1,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 \\ 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 1,0 & 0,0 & 0,0 & 0,0 & 0,0 \\ 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 1,0 & 0,3 & 0,3 & 0,3 \\ 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,3 & 1,0 & 0,3 & 0,3 \\ 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,3 & 0,3 & 1,0 & 0,3 \\ 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 & 0,3 & 0,3 & 0,3 & 1,0 \end{bmatrix}$$

$$C_1 = ZG + S', \; C_3 = ZS + R, \; V = ZGZ' + ZS + S'Z' + R$$

$$V = \begin{bmatrix}
1{,}8 & 1{,}1 & 1{,}1 & 0{,}8 & 0{,}8 & 0{,}8 & 0{,}8 & 0{,}8 & 0{,}3 & 0{,}3 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 \\
1{,}1 & 1{,}8 & 1{,}1 & 0{,}8 & 0{,}8 & 0{,}8 & 0{,}8 & 0{,}8 & 0{,}3 & 0{,}3 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 \\
1{,}1 & 1{,}1 & 1{,}8 & 0{,}8 & 0{,}8 & 0{,}8 & 0{,}8 & 0{,}8 & 0{,}3 & 0{,}3 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 \\
0{,}8 & 0{,}8 & 0{,}8 & 1{,}8 & 1{,}1 & 1{,}1 & 1{,}1 & 1{,}1 & 0{,}3 & 0{,}3 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 \\
0{,}8 & 0{,}8 & 0{,}8 & 1{,}1 & 1{,}8 & 1{,}1 & 1{,}1 & 1{,}1 & 0{,}3 & 0{,}3 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 \\
0{,}8 & 0{,}8 & 0{,}8 & 1{,}1 & 1{,}1 & 1{,}8 & 1{,}1 & 1{,}1 & 0{,}3 & 0{,}3 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 \\
0{,}8 & 0{,}8 & 0{,}8 & 1{,}1 & 1{,}1 & 1{,}1 & 1{,}8 & 1{,}1 & 0{,}3 & 0{,}3 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 \\
0{,}8 & 0{,}8 & 0{,}8 & 1{,}1 & 1{,}1 & 1{,}1 & 1{,}1 & 1{,}8 & 0{,}3 & 0{,}3 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 \\
0{,}3 & 0{,}3 & 0{,}3 & 0{,}3 & 0{,}3 & 0{,}3 & 0{,}3 & 0{,}3 & 1{,}8 & 1{,}1 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 \\
0{,}3 & 0{,}3 & 0{,}3 & 0{,}3 & 0{,}3 & 0{,}3 & 0{,}3 & 0{,}3 & 1{,}1 & 1{,}8 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 \\
0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 2{,}3 & 1{,}3 & 1{,}3 & 1{,}3 & 1{,}3 \\
0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 1{,}3 & 2{,}3 & 1{,}6 & 1{,}6 & 1{,}6 \\
0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 1{,}3 & 1{,}6 & 2{,}3 & 1{,}6 & 1{,}6 \\
0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 1{,}3 & 1{,}6 & 1{,}6 & 2{,}3 & 1{,}6 \\
0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 0{,}0 & 1{,}3 & 1{,}6 & 1{,}6 & 1{,}6 & 2{,}3
\end{bmatrix}$$

Den generelle sammensetningen på V er gitt i IV 9.

$$\Gamma = X + ZP + T$$

$$\Gamma' = \begin{bmatrix}
{,}45 & {,}45 & {,}45 & -{,}53 & -{,}53 & -{,}53 & -{,}53 & -{,}53 & {,}71 & {,}71 & {,}85 & -{,}16 & -{,}16 & -{,}16 & -{,}16 \\
{,}55 & {,}55 & {,}55 & 1{,}53 & 1{,}53 & 1{,}53 & 1{,}53 & 1{,}53 & {,}29 & {,}29 & {,}15 & 1{,}16 & 1{,}16 & 1{,}16 & 1{,}16
\end{bmatrix}$$

Estimatene for $\underline{b}$ beregnes som $(\Gamma'V^{-1}\Gamma)^{-}\Gamma'V^{-1}\underline{y}$. Således:

$$\hat{\underline{b}} = \begin{bmatrix} \hat{h}_1 \\ \hat{h}_2 \end{bmatrix} = \begin{bmatrix} 0{,}82567 & 0{,}43193 \\ 0{,}43193 & 0{,}53430 \end{bmatrix} \begin{bmatrix} -170{,}976942 \\ 326{,}153238 \end{bmatrix} = \begin{bmatrix} -0{,}294804 \\ 100{,}411500 \end{bmatrix}$$

$$\hat{\underline{u}} = P\hat{\underline{b}} + (GZ' + S)V^{-1}(\underline{y} - \Gamma\hat{\underline{b}})$$

$$\begin{bmatrix} \hat{s}_1 \\ \hat{s}_2 \\ \hat{s}_3 \end{bmatrix} = \begin{bmatrix} -0{,}50 & 0{,}50 \\ -0{,}25 & 0{,}25 \\ -0{,}10 & 0{,}10 \end{bmatrix} \begin{bmatrix} -0{,}294804 \\ 100{,}411500 \end{bmatrix} + \begin{bmatrix} 1{,}938743 \\ 0{,}510495 \\ -2{,}694161 \end{bmatrix} = \begin{bmatrix} 52{,}291895 \\ 25{,}687071 \\ 7{,}376469 \end{bmatrix}$$

Bruker en derimot "mixed-model" ligningene direkte (sml. IV 10):

$$\begin{bmatrix} 2,6250 & -1,3459 & 0,0230 & 0,6608 & 0,6548 \\ -1,3459 & 9,6747 & 3,5754 & 1,7462 & 3,3078 \\ 0,0230 & 3,5754 & 2,9502 & 1,5482 & 0,0 \\ 0,6608 & 1,7462 & 1,5482 & 1,7588 & 0,0 \\ 0,6548 & 3,3078 & 0,0 & 0,0 & 5,2590 \end{bmatrix} \begin{bmatrix} \hat{h}_1 \\ \hat{h}_2 \\ \hat{t}_1 \\ \hat{t}_2 \\ \hat{t}_3 \end{bmatrix} = \begin{bmatrix} -138,4653 \\ 973,5383 \\ 368,4632 \\ 179,1472 \\ 316,2090 \end{bmatrix}$$

Den inverse koeffisientmatrisen er:

$$\begin{bmatrix} 0,82567 & 0,43193 & -0,26406 & -0,50660 & -0,37450 \\ 0,43193 & 0,53430 & -0,53404 & -0,22266 & -0,38985 \\ -0,26406 & -0,53404 & 1,22273 & -0,44688 & 0,36879 \\ -0,50660 & -0,22266 & -0,44688 & 1,37332 & 0,20313 \\ -0,37450 & -0,38985 & 0,36879 & 0,20313 & 0,48200 \end{bmatrix}$$

og lösningene:

$$\begin{bmatrix} \hat{h}_1 \\ \hat{h}_2 \\ \hat{t}_1 \\ \hat{t}_2 \\ \hat{t}_3 \end{bmatrix} = \begin{bmatrix} -0,294804 \\ 100,411500 \\ 3,741101 \\ -1,019726 \\ -2,993513 \end{bmatrix}$$

$$G\hat{\underline{t}} = \begin{bmatrix} 0,6 & 0,3 & 0,0 \\ 0,3 & 0,6 & 0,0 \\ 0,0 & 0,0 & 0,9 \end{bmatrix} \begin{bmatrix} 3,741101 \\ -1,019726 \\ -2,993513 \end{bmatrix} = \begin{bmatrix} 1,938743 \\ 0,510495 \\ -2,694161 \end{bmatrix}$$

$$= (GZ' + S)V^{-1}(\underline{y} - \Gamma\hat{\underline{b}})$$

$\hat{\underline{u}} = G\hat{\underline{t}} + P\hat{\underline{b}}$, og $\hat{\underline{u}}$ er vektoren for avlsverdiene på fedrene.

EKSEMPEL I. HENSYN TAS TIL PERMANENT MILJÖEFFEKT (HENDERSON, 1972)

Vi går ut fra fölgende datasett:

		Besetning-år (hy)				
Gruppe (g)	Far (s)	11	12	13	24	25
1	1	14, 12			12, 16	
1	2	15	18, 12	16		
1	3		20		13, 22	
2	4	15			17, 15	13
2	5				18	14

Vi forutsetter först at den permanente miljöeffekten (C^2) influerer på döttre i samme besetning-år kombinasjon, med andre ord, et samspill mellom far og besetning-år. Modellen blir da:

$$Y_{lijkm} = (hy)_{jk} + g_l + s_{li} + (shy)_{i,jk} + e_{lijkm}$$

g og hy er "fixed" effekter.

Videre settes:

$$\frac{\sigma_e^2}{\sigma_s^2} = 15 \quad \text{og} \quad \frac{\sigma_e^2}{\sigma_{shy}^2} = 30$$

Noen av BLUP-ligningene er vist nedenfor:

Gruppe 1: $11\ g_1 + 4\ s_1 + 4\ s_2 + 3\ s_3 + 3\ hy_{11} + 3\ hy_{12} + hy_{13} + 4\ hy_{24} + 2\ shy_{1,11} + 2\ shy_{1,24} + shy_{2,11} + 2\ shy_{2,12} + shy_{2,13} + shy_{3,12} + 2\ shy_{3,24} = 170$

Far 1: $4\ g_1 + 19\ s_1 + 2\ hy_{11} + 2\ hy_{24} + 2\ hy_{1,11} + 2\ shy_{1,24} = 54$

hy_{11}: $3\ g_1 + g_2 + 2\ s_1 + s_2 + s_4 + 4\ hy_{11} + 2\ shy_{1,11} + shy_{2,11} + shy_{4,11} = 56$

$shy_{1,11}$: $2\ g_1 + 2\ s_1 + 2\ hy_{11} + {}^*32\ shy_{1,11} = 26$

$^*2 + 30 = 32$

En lösning på ligningene er ["constraint" $\hat{g}_1 = 0$ fordi (hy)-ligningene summerer opp til g-ligningene].

$\hat{g}' = [-1,746 \quad 0]$

$\hat{s}' = [-0,210 \quad -0,195 \quad 0,405 \quad -0,080 \quad 0,080]$

$\hat{hy}' = [15,492 \quad 18,439 \quad 19,511 \quad 17,100 \quad 13,500]$

og estimatene på shy-verdiene er vist i tabellen nedenfor:

	Besetning-år				
Far	11	12	13	24	25
1	-0,033			-0,071	
2	0,047	-0,094	-0,051		
3		0,094		0,109	
4	-0,013		0,051	-0,064	-0,014
5				0,026	0,014

Observér at: $\sum_{jk} \hat{shy}_{ijk} = \frac{15}{30} \hat{s}_i$

$\sum_{i} \hat{shy}_{ijk} = 0$

Det kan også tenkes at den permanente miljöeffekten (C^2) influerer på dötre etter en far i samme besetning (ikke besetning-år). Da blir modellen:

$$Y_{lijkm} = (hy)_{jk} + g_l + s_{li} + (sh)_{ij} + e_{lijkm}$$

og noen av BLUP-ligningene $[\frac{\sigma_e^2}{\sigma_s^2} = 15 \; ; \; \frac{\sigma_e^2}{\sigma_{sh}^2} = 30]$ er som fölger:

Gruppe 1: $11\ g_1 + 4\ s_1 + 4\ s_2 + 3\ s_3 + 3\ hy_{11} + 3\ hy_{12} + hy_{13} + 4\ hy_{24} + 2\ sh_{11} + 2\ sh_{12} + 4\ sh_{21} + sh_{31} + 2\ sh_{32} = 170$

Far 1: $4\ g_1 + 19\ s_1 + 2\ hy_{11} + 2\ hy_{24} + 2\ sh_{11} + 2\ sh_{12} = 54$

hy_{11}: $3\ g_1 + g_2 + 2\ s_1 + s_2 + s_4 + 4\ hy_{11} + 2\ sh_{11} + sh_{21} + sh_{41} = 56$

sh_{21}: $4\ g_1 + 4\ s_2 + hy_{11} + 2\ hy_{12} + hy_{13} + 34\ sh_{21} = 61$

En lösning på ligningene er [$\hat{g}_1 = 0$, "constraint"]:

$$\hat{g}' = [-1,716 \quad 0]$$

$$\hat{s}' = [-0,215 \quad -0,189 \quad 0,404 \quad -0,078 \quad 0,078]$$

$$\hat{hy}' = [15,493 \quad 18,405 \quad 19,520 \quad 17,086 \quad 13,518]$$

og tabell over $\hat{sh}$:

	Besetning	
Far	1	2
1	-0,035	-0,072
2	-0,094	-
3	0,094	0,108
4	0,035	-0,074
5	-	0,039

Legg merke til at: $\sum_j \hat{sh}_{ij} = \frac{15}{30} \hat{s}_i$

$$\sum_i \hat{sh}_{ij} = 0$$

EKSEMPEL J. ESTIMERING AV AVLSVERDIEN TIL OKSER MED KORRIGERING FOR EVENTUELL SELEKSJON BLANT MÖDRENE TIL TESTDÖTRENE (HENDERSON, 1973)

Foreligger det seleksjon blant mödrene til testdötrene, kan dette före til feilaktig rangering mellom oksene. En måte å redusere denne "bias" på er å inkludere mor-fedre i modellen. Det vil si:

$$y_{ijklmn} = \mu + h_i + \frac{1}{2}(g_j + s_{jk}) + g_l + s_{lm} + e_{ijklmn}$$

y_{ijklmn} = registrering på n^{te} datter etter m^{te} far i l^{te} fargruppe, k^{te} morfar i j^{te} morfargruppe og i^{te} besetning-år-sesong. Det er altså forutsatt bare <u>en registrering</u> pr. datter

h_i = effekten av i^{te} besetning-år-sesong ("fixed")

g_j = effekten av j^{te} morfar-gruppe ("fixed")

s_{jk} = effekten av k^{te} morfar innen j^{te} morfargruppe ("random") osv.

Skriver vi denne modellen ved bruk av matrisealgebra, får vi:

$$\underline{y} = X\underline{h} + M\underline{g} + Z\underline{s} + \underline{e} \text{ , hvor}$$

$$\underline{h} = \begin{bmatrix} \mu + h_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} \text{, osv. (se forövrig side V J:3)}$$

Fölgelig blir "mixed-model"-ligningene:

$$\begin{bmatrix} X'X & X'M & X'Z \\ M'X & M'M & M'Z \\ Z'X & Z'M & Z'Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\underline{h}} \\ \hat{\underline{g}} \\ \hat{\underline{s}} \end{bmatrix} = \begin{bmatrix} X'\underline{y} \\ M'\underline{y} \\ Z'\underline{y} \end{bmatrix}$$

M'M og Z'Z er ikke diagonalmatriser.

$$G^{-1} = \frac{\sigma_e^2}{\sigma_s^2}$$. Som vanlig er $\sigma_s^2 = \frac{1}{4}h^2\sigma_y^2$.

Derimot er $\sigma_e^2 = (1 - \frac{1}{16}h^2 - \frac{1}{4}h^2)\sigma_y^2 = (1 - \frac{5}{16}h^2)\sigma_y^2$.

Arbeider en f.eks. med flere laktasjoner pr. datter, må en dele opp σ_e^2 på fölgende måte:

$$\sigma_e^2 = (r - \frac{5}{16}h^2)\sigma_y^2 + (1 - r)\sigma_y^2 \quad , \text{ hvor}$$

r = reproduserbarheten.

Vi skal illustrere metoden med et numerisk eksempel:

	Besetning 1				Besetning 2			
Mor-far	Far 1	5	6	7	Far 1	5	6	7
1	-	12	21;15;9	-	-	13;18	11	-
2	-	12	-	18	-	15;14;17	-	22;12
3	13;20;15;17	-	-	15	19	-	13;12	14
4	16;19	10;12	-	-	-	13	15;14	-

I dette materialet har hver datter bare en registrering (laktasjon). Om $h^2 = \frac{16}{41}$, blir: $\sigma_e^2 = (1 - \frac{5}{16}h^2)\sigma_y^2 = \frac{36}{41}\sigma_y^2$.

$$\sigma_s^2 = \frac{1}{4}h^2\sigma_y^2 = \frac{4}{41}\sigma_y^2 \quad , \text{ og fölgelig } G^{-1} = 9, \text{ som da skal}$$

adderes til diagonalelementene i far- og morfarligningene.

Vi forutsetter at:

$$g_1 = \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}, \qquad g_2 = \begin{bmatrix} s_3 \\ s_4 \end{bmatrix}, \qquad g_3 = \begin{bmatrix} s_5 \\ s_6 \\ s_7 \end{bmatrix}$$

Nedenfor har en for dette eksemplet satt opp de ulike leddene i modellen for en del av observasjonene:

$$\underline{y} = \quad X \quad \underline{b} \quad M \quad \underline{g} \quad Z \quad \underline{s} + \underline{e}$$

$$\begin{bmatrix} y_{111351} \\ y_{111361} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ y_{223111} \\ \vdots \end{bmatrix}
\begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ \vdots & \end{bmatrix}
\begin{bmatrix} \mu + h_1 \\ \mu + h_2 \end{bmatrix}
+
\begin{bmatrix} \frac{1}{2} & 0 & 1 \\ \frac{1}{2} & 0 & 1 \\ \frac{1}{2} & 0 & 1 \\ \frac{1}{2} & 0 & 1 \\ \frac{1}{2} & 0 & 1 \\ \frac{1}{2} & 0 & 1 \\ \frac{1}{2} & 0 & 1 \\ \frac{1}{2} & 0 & 1 \\ \frac{1}{2} & 0 & 1 \\ \frac{1}{2} & 0 & 1 \\ \frac{1}{2} & 0 & 1 \\ \frac{1}{2} & 0 & 1 \\ \frac{1}{2} & 0 & 1 \\ \frac{1}{2} & 0 & 1 \\ 1 & \frac{1}{2} & 0 \\ 1 & \frac{1}{2} & 0 \\ 1 & \frac{1}{2} & 0 \\ 1 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 1 \\ 1 & \frac{1}{2} & 0 \\ & \vdots & \end{bmatrix}
\begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix}
+
\begin{bmatrix} \frac{1}{2} & 0 & 0 & 0 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 & 1 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 & 1 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 & 1 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 1 \\ 0 & \frac{1}{2} & 0 & 0 & 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 1 \\ 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 1 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 1 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 1 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 1 \\ 1 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ & & & \vdots & & & \end{bmatrix}
\begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \\ s_6 \\ s_7 \end{bmatrix}$$

Noen av "mixed model"-ligningene er vist nedenfor:

$$g_1 : 10{,}5\ g_1 + 3{,}5\ g_2 + 7\ g_3 + 8{,}75\ s_1 + 1{,}75\ s_2 + 2{,}5\ s_3 +$$
$$s_4 + 3{,}5\ s_5 + 2\ s_6 + 1{,}5\ s_7 + 9\ h_1 + 5\ h_2 = 223{,}50$$

$$s_2 : 8{,}75\ g_1 + 3{,}5\ g_2 + 3{,}5\ g_3 + 17{,}7\ s_1 + 2{,}5\ s_3 + s_4 +$$
$$1{,}5\ s_5 + 2\ s_6 + 8\ h_1 + 2{,}5\ h_2 = 168{,}50$$

$$h_2 : 5\ g_1 + 3{,}5\ g_2 + 14\ g_3 + 2{,}5\ s_1 + 2{,}5\ s_2 + 2\ s_3 + 1{,}5\ s_4 +$$
$$6\ s_5 + 5\ s_6 + 3\ s_7 + 15\ h_2 = 222$$

Om en setter $\hat{h}_1 = 0$ (dvs. 1 "constraint") blir:

$$\hat{g}' = [13{,}391 \quad 9{,}420 \quad 8{,}851]$$

$$\hat{s}' = [\underbrace{-0{,}155 \quad 0{,}155}_{0} \quad \underbrace{-0{,}068 \quad 0{,}068}_{0} \quad \underbrace{-0{,}492 \quad -0{,}168 \quad 0{,}660}_{0}]$$

$$\hat{h}' = [-1{,}087 \quad 0]$$

En kan merke seg her at $\Sigma\hat{s} = 0$ innen hver gruppe. Videre gir $\hat{g} + \hat{s}$ avlsverdien til den enkelte okse.

EKSEMPEL K. RANGERING AV INDIVIDER NÅR EN BRUKER FLERE REGISTRERINGER PÅ DET ENKELTE INDIVID (HENDERSON, 1973)

Et datasett kan ha fölgende struktur:

	Periode (p)		
Ku (c)	1	2	3
1	15	13	19
2	10	20	-
3	-	14	16
4	-	17	-
5	-	-	12

En kan da reise fölgende spörsmål: Hva er RPA ("real producing ability") og ETA ("estimated transmitting ability") for disse kyrne?

Modellen kan skrives som:

$$\underline{y} = X\underline{b} + Z\underline{c} + \underline{e}$$

$$\underline{b} = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} \qquad \text{og } \underline{c} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \end{bmatrix}$$

$$X = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$Z = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$X'X = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

$$X'Z = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

$$Z'Z = \begin{bmatrix} 3 & & & & \\ & 2 & & & \\ & & 2 & & 0 \\ 0 & & & 1 & \\ & & & & 1 \end{bmatrix} \qquad X'y = \begin{bmatrix} 25 \\ 64 \\ 47 \end{bmatrix} \qquad Z'y = \begin{bmatrix} 47 \\ 30 \\ 30 \\ 17 \\ 12 \end{bmatrix}$$

Vi antar at de 5 kyrne er beslektet på fölgende måte:

$$A = \begin{bmatrix} 1 & \frac{1}{4} & \frac{1}{4} & \frac{1}{16} & \frac{1}{16} \\ \frac{1}{4} & 1 & \frac{1}{4} & \frac{1}{16} & \frac{1}{16} \\ \frac{1}{4} & \frac{1}{4} & 1 & \frac{1}{16} & \frac{1}{16} \\ \frac{1}{16} & \frac{1}{16} & \frac{1}{16} & 1 & \frac{1}{4} \\ \frac{1}{16} & \frac{1}{16} & \frac{1}{16} & \frac{1}{4} & 1 \end{bmatrix}$$

Dette innebärer at: G inneholder $\frac{h^2 a_{ii} + (r - h^2) I}{(1 - r)}$

som diagonalelement og $\frac{h^2 a_{ij}}{(1 - r)}$ som "off-diagonal"-element.

a_{ij} er elementene i A.

Med $h^2 = 0{,}25$ og reproduserbarheten = $r = 0{,}5$ blir:

$$G = \begin{bmatrix} 1 & \frac{1}{8} & \frac{1}{8} & \frac{1}{32} & \frac{1}{32} \\ \frac{1}{8} & 1 & \frac{1}{8} & \frac{1}{32} & \frac{1}{32} \\ \frac{1}{8} & \frac{1}{8} & 1 & \frac{1}{32} & \frac{1}{32} \\ \frac{1}{32} & \frac{1}{32} & \frac{1}{32} & 1 & \frac{1}{8} \\ \frac{1}{32} & \frac{1}{32} & \frac{1}{32} & \frac{1}{8} & 1 \end{bmatrix}$$

Som en lett kan se, forutsetter en her at registreringene har samme varians, og at det additive genetiske slektskapet er den eneste årsak til kovarians mellom beslektede individer.

"Mixed-model"-ligningene blir:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\underline{b}} \\ \hat{\underline{c}} \end{bmatrix} = \begin{bmatrix} X'\underline{y} \\ Z'\underline{y} \end{bmatrix}$$, eller skrevet om:

$$\begin{bmatrix} X'X & X'ZG \\ GZ'X & GZ'ZG + G \end{bmatrix} \begin{bmatrix} \hat{\underline{b}} \\ \hat{\underline{v}} \end{bmatrix} = \begin{bmatrix} X'\underline{y} \\ GZ'\underline{y} \end{bmatrix}$$, hvor

$\hat{\underline{v}} = G^{-1}\hat{\underline{c}}$

$\hat{\underline{c}}$ = RPA og $\hat{\text{ETA}} = \frac{h^2}{1 - r} A \hat{\underline{v}}$ for individer <u>med</u> registreringer.

En kan selvsagt også beregne ETA for individer som <u>selv ikke</u> har registreringer, men som er beslektet med individer som har registreringer. Da blir:

$$\hat{ETA} = \frac{h^2}{1 - r} A_{12}\hat{\underline{v}}$$

A_{12} = additive genetiske slektskapet mellom individer som selv har registreringer (2) og individer som ikke har registreringer (1)

En skal ikke gå noe närmere inn på selve beregningene her.

Oppgave:

Beregn RPA og ETA for den enkelte ku i eksemplet foran.

Svar:

$$\hat{\underline{v}}' = [0,959 \quad 1,009 \quad -0,685 \quad 0,882 \quad -2,164]$$

$$\hat{ETA} = \frac{1}{2} A \hat{\underline{v}}$$

VI. BLUE OG BLUP I EN SELEKSJONSMODELL

1. Teori

"Mixed linear models" har relativt lenge värt akseptert som et nyttig hjelpemiddel ved analysering av data i husdyravlen. Med andre ord en har hatt mulighet til å beregne BLUE av estimērbare lineäre funksjoner av "fixed" elementene i modellen, videre BLUP av "random" elementene i modellen. Et stort problem ved data i husdyravlen er at de kan väre påvirket av seleksjon, noe som kan medföre "bias" i estimatet av "fixed" effekt og i prediksjonen av "random" effekt. En skal i det fölgende se hvordan en slik situasjon best kan löses. Det er först og fremst Henderson (1968, 1975 a) som har bidratt til å löse dette problemet. Problemet er imidlertid diskutert også av andre (f.eks. Rönningen, 1971; Fimland, 1975b, 1976 b).

La oss ta utgangspunkt i den vanlige "mixed-model":

$$\underline{y} = X\underline{b} + Z\underline{u} + \underline{e}$$

Vi tenker oss en ny vektor $\underline{w}$, som vi definerer som: $w = \mu_w + k$ hvor k er angitt som avvik fra μ_w. k er altså "selection bias" og $\underline{w}$ kan väre korrelert med $\underline{y}$, $\underline{u}$ eller $\underline{e}$.

Varians- og kovariansmatrisen for $\underline{y}$, $\underline{u}$, $\underline{e}$ og $\underline{w}$ setter vi lik:

$$\text{Var.-Kov.} \begin{bmatrix} \underline{y} \\ \underline{u} \\ \underline{e} \\ \underline{w} \end{bmatrix} = \begin{bmatrix} V & ZG & R & ZD+K \\ GZ' & G & O & D \\ R & O & R & T \\ D'Z'+K' & D' & T' & H \end{bmatrix} \sigma^2$$

En går ut fra at G og R er symmetriske.

Videre er $V\sigma^2$ som tidligere vist lik:

$$(ZGZ' + R)\ \sigma^2$$

Det kan bevises (Henderson, 1968, 1975 a ; Rönningen, 1971; Fimland, 1975 b) at $\underline{\hat{k} = H\hat{\phi}}$ hvor $\hat{\underline{\phi}}$ finnes ut fra fölgende BLUE-ligninger:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}T & X'R^{-1}Z \\ T'R^{-1}X & T'R^{-1}T+D'G^{-1}D & T'R^{-1}Z-D'G^{-1} \\ Z'R^{-1}X & Z'R^{-1}T-G^{-1}D & Z'R^{-1}Z+G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\underline{b}} \\ \hat{\underline{\phi}} \\ \hat{\underline{u}} \end{bmatrix} = \begin{bmatrix} X'R^{-1}\underline{y} \\ T'R^{-1}\underline{y} \\ Z'R^{-1}\underline{y} \end{bmatrix}$$

Disse kan utledes ved bruk av "LaGrange multipliers" (Henderson,1968, 1975 a) eller ved bruk av "maximum likelihood" (Fimland, 1975 b), og er da de modifiserte "mixed model" ligningene.

2. Spesialtilfeller

I det fölgende vil en benytte 2 matriser (evt. vektorer), henholdsvis L' og M'. Som en vil komme inn på i noen av eksemplene senere, definérer en disse matrisene eller vektorene alt etter problemstillingen.

a. Seleksjon på y

Definér $w = L'\underline{y}$. Derfor:

$$D = \frac{1}{\sigma^2}\,\text{kov}(\underline{u}\underline{w}') = \frac{1}{\sigma^2}\,\text{kov}(\underline{u}\underline{y}'L) = GZ'L$$

$$T = \frac{1}{\sigma^2}\,\text{kov}(\underline{e}\underline{w}') = \frac{1}{\sigma^2}\,\text{kov}(\underline{e}\underline{y}'L) = RL$$

Ved å benytte dette blir BLUE-ligningene:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}RL & X'R^{-1}Z \\ L'RR^{-1}X & L'RR^{-1}RL+L'ZGG^{-1}GZ'L & L'RR^{-1}Z-L'ZGG^{-1} \\ Z'R^{-1}X & Z'R^{-1}RL-G^{-1}GZ'L & Z'R^{-1}Z+G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\underline{b}} \\ \hat{\underline{\phi}} \\ \hat{\underline{u}} \end{bmatrix} = \begin{bmatrix} X'R^{-1}\underline{y} \\ L'RR^{-1}\underline{y} \\ Z'R^{-1}\underline{y} \end{bmatrix}$$

Fra tidligere vet vi at $V = R + ZGZ'$. Derfor,

$$\begin{bmatrix} X'R^{-1}X & X'L & X'R^{-1}Z \\ L'X & L'VL & 0 \\ Z'R^{-1}X & 0 & Z'R^{-1}Z+G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\underline{b}} \\ \hat{\underline{\phi}} \\ \hat{\underline{u}} \end{bmatrix} = \begin{bmatrix} X'R^{-1}\underline{y} \\ L'\underline{y} \\ Z'R^{-1}\underline{y} \end{bmatrix}$$

En kan her legge merke til at med $L'X = 0$ vil $\hat{\underline{\phi}}$ ikke ha noen innflytelse på $\hat{\underline{b}}$ og $\hat{\underline{u}}$. $H = L'VL$.

b. Seleksjon på e

Definér $\underline{w} = M'\underline{e}$. Derfor,

$$D = \frac{1}{\sigma^2}\ \mathrm{kov}(\underline{u}\underline{e}'M) = 0$$

$$T = \frac{1}{\sigma^2}\ \mathrm{kov}(\underline{e}\underline{e}'M) = RM,$$ og BLUE-ligningene:

$$\begin{bmatrix} X'R^{-1}X & X'M & X'R^{-1}Z \\ M'X & M'RM & M'Z \\ Z'R^{-1}X & Z'M & Z'R^{-1}Z+G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\underline{b}} \\ \hat{\underline{\phi}} \\ \hat{\underline{u}} \end{bmatrix} = \begin{bmatrix} X'R^{-1}\underline{y} \\ M'\underline{y} \\ Z'R^{-1}\underline{y} \end{bmatrix}$$

I dette tilfellet må både $M'X = 0$ og $M'Z = 0$ forat $\hat{\underline{\phi}}$ ikke skal ha noen innflytelse på $\hat{\underline{b}}$ og $\hat{\underline{u}}$. $H = M'RM$

c. Seleksjon på u

Definér $\underline{w} = L'\underline{u}$. Derfor,

$$D = \frac{1}{\sigma^2}\ \mathrm{kov}(\underline{u}\underline{u}'L) = GL$$

$$T = \frac{1}{\sigma^2}\ \mathrm{kov}(\underline{e}\underline{u}'L) = 0,$$ og BLUE-ligningene:

$$\begin{bmatrix} X'R^{-1}X & 0 & X'R^{-1}Z \\ 0 & L'GL & -L' \\ Z'R^{-1}X & -L & Z'R^{-1}Z+G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\underline{b}} \\ \hat{\underline{\phi}} \\ \hat{\underline{u}} \end{bmatrix} = \begin{bmatrix} X'R^{-1}\underline{y} \\ 0 \\ Z'R^{-1}\underline{y} \end{bmatrix}$$

$H = L'GL$

d. Seleksjon på $\underline{u}$ og $\underline{e}$ hver for seg

$$\underline{w} = \begin{bmatrix} L'\underline{u} \\ M'\underline{e} \end{bmatrix}, \quad \underline{\phi} = \begin{bmatrix} \underline{\phi}_u \\ \underline{\phi}_e \end{bmatrix}$$

Derfor,

$$D = \frac{1}{\sigma^2} \text{ kov}[\underset{-}{u}(\underset{-}{u}'L \quad \underset{-}{e}'M)] = [GL \quad 0]$$

$$T = \frac{1}{\sigma^2} \text{ kov}[\underset{-}{e}(\underset{-}{u}'L \quad \underset{-}{e}'M)] = [0 \quad RM]$$

Videre,

$$X'R^{-1}T = X'R^{-1}[0 \quad RM] = [0 \quad X'M]$$

$$T'R^{-1}T = \begin{bmatrix} 0 \\ M'R \end{bmatrix} R^{-1}[(0 \quad RM)] = \begin{bmatrix} 0 & 0 \\ 0 & M'RM \end{bmatrix}$$

$$D'G^{-1}D = \begin{bmatrix} L'G \\ 0 \end{bmatrix} G^{-1}[GL \quad 0] = \begin{bmatrix} L'GL & 0 \\ 0 & 0 \end{bmatrix}$$

$$T'R^{-1}Z = \begin{bmatrix} 0 \\ M'R \end{bmatrix} R^{-1}Z = \begin{bmatrix} 0 \\ M'Z \end{bmatrix}$$

$$D'G^{-1} = \begin{bmatrix} L'G \\ 0 \end{bmatrix} G^{-1} = \begin{bmatrix} L' \\ 0 \end{bmatrix}$$

$$T'R^{-1}\underset{-}{y} = \begin{bmatrix} 0 \\ M'R \end{bmatrix} R^{-1}\underset{-}{y} = \begin{bmatrix} 0 \\ M'\underset{-}{y} \end{bmatrix}$$

og BLUE-ligningene:

$$\begin{bmatrix} X'R^{-1}X & 0 & X'M & X'R^{-1}Z \\ 0 & L'GL & 0 & -L' \\ M'X & 0 & M'RM & M'Z \\ Z'R^{-1}X & -L & Z'M & Z'R^{-1}Z+G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\underset{-}{b}} \\ \hat{\underset{-}{\phi}}_u \\ \hat{\underset{-}{\phi}}_e \\ \hat{\underset{-}{u}} \end{bmatrix} = \begin{bmatrix} X'R^{-1}\underset{-}{y} \\ 0 \\ M'\underset{-}{y} \\ Z'R^{-1}\underset{-}{y} \end{bmatrix}$$

og

$$H = \begin{bmatrix} L'GL & 0 \\ 0 & M'RM \end{bmatrix}$$

Kjennskap til $\underline{H}$ er viktig for beregning av $\underline{k}$:

$$\hat{\underline{\phi}} = H^{-1}\hat{\underline{k}} \quad \text{eller} \quad \hat{\underline{k}} = H\hat{\underline{\phi}}$$

Hva er så variansen på $\hat{k}$? I fölge Henderson (1968) finner vi denne som vist nedenfor:

$$\begin{aligned}
\text{Var}\ (\hat{\underline{k}}) &= E(\hat{\underline{k}}\hat{\underline{k}}') - (E\hat{\underline{k}})^2 = \\
&E(H\hat{\underline{\phi}}\hat{\underline{\phi}}'H') = H\ \text{Var}\ (\hat{\underline{\phi}})\ H' = \\
&H(U - H^{-1})H'\ \sigma^2 = \underline{(HUH' - H')\sigma^2}, \quad (H = H')
\end{aligned}$$

hvor U kommer fra:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}T & X'R^{-1}Z \\ T'R^{-1}X & T'R^{-1}T+G'G^{-1}D & T'R^{-1}Z-D'G^{-1} \\ Z'R^{-1}X & Z'R^{-1}T-G^{-1}D & Z'R^{-1}Z+G^{-1} \end{bmatrix}^{-1} = \begin{bmatrix} P & W & Q \\ W' & U & V \\ Q' & V' & S \end{bmatrix}$$

Det kan tilföyes her at ifölge Fimland (1976 a) er Var. $(\hat{\underline{k}}) = HUH'\ \sigma^2$, altså et resultat som avviker fra Henderson (1968).

Fimland (1976 a) forutsetter i sitt resonnement at variansen på $\underline{w}$ er den samme i selektert "sample" og i uselektert populasjon (dvs. $H = H_s$).

3. Betingete gjennomsnittstall i seleksjonstilfeller

Om vi kan forutsette normalfordeling, blir det betingete gjennomsnittstall for henholdsvis $\underline{u}$, $\underline{y}$ og $\underline{e}$ når det er selektert på $\underline{y}$ lik:

$$E(\underline{u}|L'\underline{y}) = GZ'L(L'VL)^{-1}\overbrace{(L'\underline{y} - L'X\underline{b})}^{\text{angir avvik fra forventningen i uselektert tilfelle}}$$

$$\begin{aligned}
E(\underline{y}|L'\underline{y}) &= X\underline{b} + VL(L'VL)^{-1}(L'\underline{y} - L'X\underline{b}) = \\
&X\underline{b} + VL\underline{\phi} \quad \text{ettersom} \quad \underline{\phi} = (L'VL)^{-1}(L'\underline{y} - L'X\underline{b})
\end{aligned}$$

$$E(\underline{e}|L'\underline{y}) = RL(L'VL)^{-1}(L'\underline{y} - L'X\underline{b})$$

Se bl.a. eksempel 4 c for illustrasjon.

4. Eksempler

a. La oss tenke oss 2 "treatments", videre 3 fedre som har det nedenfor gitte antall avkom (n_{ij}) [Rönningen (1971)]:

Treatments (t_i)	Fedre (s_j) 1	2	3	Sum
1	5	5	5	15
2	10	8	0	18
Sum	15	13	5	33

Sum ytelse er:

	1	2	3
1	40	45	30
2	119	97	–

Denne modellen som ligger bak dette materialet er:

$$y = X\underline{b} + Z\underline{u} + \underline{e} \text{ , hvor}$$

$\underline{b}' = [\mu \quad t_1 \quad t_2]$

$\underline{u}' = [s_1 \quad s_2 \quad s_3]$

En forutsetter ikke noe samspill

s: NID(0, $\frac{1}{5}\sigma^2$) ; e: NID(0, σ^2) ; t_i: fixed

$$X = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \quad \text{og } Z = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

$E(\underline{\underline{e}}\underline{\underline{e}}') = R\sigma^2 = I_{33}\sigma^2$

$$E(\underline{\underline{u}}\underline{\underline{u}}') = G\sigma^2 = \begin{bmatrix} \frac{1}{5} & 0 & 0 \\ 0 & \frac{1}{5} & 0 \\ 0 & 0 & \frac{1}{5} \end{bmatrix} \sigma^2$$

Seleksjon på y: Dette kan f.eks. innebäre en mistanke om at fedrene 1 og 2 har avkom i trt. 2 fordi deres avkom gjorde det bedre i trt. 1 sammenlignet med avkommet til far 3.

Det förste en bör undersöke er da L'X. I dette tilfellet blir L':

$$L' = [\tfrac{1}{10}\ \tfrac{1}{10}\ \tfrac{1}{10}\ \tfrac{1}{10}\ \tfrac{1}{10}\ \tfrac{1}{10}\ \tfrac{1}{10}\ \tfrac{1}{10}\ \tfrac{1}{10}\ \tfrac{1}{10}\ - \tfrac{1}{5} - \tfrac{1}{5} - \tfrac{1}{5} - \tfrac{1}{5} - \tfrac{1}{5}\ \underbrace{0 \ \ldots\ldots\ 0}_{\text{ialt } 18}]$$

og L'X = 0. Dette betyr at $\hat{\underline{\underline{\phi}}}$ ikke har noen innvirkning på $\hat{\underline{\underline{b}}}$ og $\hat{\underline{\underline{u}}}$ (dvs. intra-fixed effekt seleksjon). Både s_j og $t_2 - t_1$ er estimérbare.

Seleksjon på u: Her antar en f.eks. at far nr. 1 er bedre enn de 2 andre fedrene. Denne antagelsen er basert på informasjon om slektningene til de 3 fedrene.

Her får en da (opprinnelig eksempel):

$$L'\underline{\underline{u}} = [1\ -\tfrac{1}{2}\ -\tfrac{1}{2}] \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix}$$

Videre,

$$L'GL = [1\ -\tfrac{1}{2}\ -\tfrac{1}{2}] \begin{bmatrix} \frac{1}{5} & 0 & 0 \\ 0 & \frac{1}{5} & 0 \\ 0 & 0 & \frac{1}{5} \end{bmatrix} \begin{bmatrix} 1 \\ -\frac{1}{2} \\ -\frac{1}{2} \end{bmatrix} = \frac{6}{20}$$

og BLUE-ligningene:

$$\left[\begin{array}{ccc|c|ccc} 33 & 15 & 18 & 0 & 15 & 13 & 5 \\ 15 & 5 & 0 & 0 & 5 & 5 & 5 \\ 18 & 0 & 18 & 0 & 10 & 8 & 0 \\ \hline 0 & 0 & 0 & 0{,}30 & -1 & 0{,}5 & 0{,}5 \\ \hline 15 & 5 & 10 & -1 & 20 & 0 & 0 \\ 13 & 5 & 8 & 0{,}5 & 0 & 18 & 0 \\ 5 & 5 & 0 & 0{,}5 & 0 & 0 & 10 \end{array}\right] \left[\begin{array}{c} \hat{\mu} \\ \hat{t}_1 \\ \hat{t}_2 \\ \hline \hat{\phi} \\ \hline \hat{s}_1 \\ \hat{s}_2 \\ \hat{s}_3 \end{array}\right] = \left[\begin{array}{c} 331 \\ 115 \\ 216 \\ \hline 0 \\ \hline 159 \\ 142 \\ 30 \end{array}\right]$$

Lösningen for ϕ blir; $\hat{\phi}$ = $\underline{1,71}$, når en bruker "constraint" $\hat{\mu} = 0$.

$$\hat{k} = H\hat{\phi} = [\,1 \;\; -\tfrac{1}{2} \;\; -\tfrac{1}{2}\,] \begin{bmatrix} \frac{1}{5} & 0 & 0 \\ 0 & \frac{1}{5} & 0 \\ 0 & 0 & \frac{1}{5} \end{bmatrix} \begin{bmatrix} 1 \\ -\frac{1}{2} \\ -\frac{1}{2} \end{bmatrix} 1{,}71 = \underline{0{,}171}$$

Det er med andre ord en indikasjon på at det her har foregått en inter-fedre (inter-genotyper) seleksjon. Signifikansen må avgjöres med en test.

Både $\hat{t}_1 - \hat{t}_2$ og s_j er estimérbare.

<u>Seleksjon på e</u>: Her antar en at e'ene tilhörende "treatment 2" har middeltall större enn 0. Således:

$$\underline{w} = M'\underline{e}\,[-\tfrac{1}{15} \;\; -\tfrac{1}{15} \;\ldots\ldots\; -\tfrac{1}{15} \;\; \tfrac{1}{18} \;\; \tfrac{1}{18} \;\ldots\ldots\; \tfrac{1}{18}] \begin{bmatrix} e_{111} \\ e_{112} \\ \cdot \\ \cdot \\ \cdot \\ e_{228} \end{bmatrix}$$

M' kunne kanskje også ha hatt fölgende struktur:

$$M' = [\underbrace{0 \;\ldots\ldots\; 0}_{15}]\,[\underbrace{1 \;\ldots\ldots\; 1}_{18}]$$

En vil imidlertid foretrekke den förstnevnte.

$M'X = [0 \quad -1 \quad 1]$

$M'Z = [0{,}2222 \quad 0{,}1111 \quad -0{,}3333]$

$M'RM = 0{,}1222$, og

BLUE-ligningene blir:

$$\left[\begin{array}{ccc|c|ccc} 33 & 15 & 18 & 0 & 15 & 13 & 5 \\ 15 & 15 & 0 & -1 & 5 & 5 & 5 \\ 18 & 0 & 18 & 1 & 10 & 8 & 0 \\ \hline 0 & -1 & 1 & 0{,}12 & 0{,}22 & 0{,}11 & -0{,}33 \\ \hline 15 & 5 & 10 & 0{,}22 & 20 & 0 & 0 \\ 13 & 5 & 8 & 0{,}11 & 0 & 18 & 0 \\ 5 & 5 & 0 & -0{,}33 & 0 & 0 & 10 \end{array}\right] \left[\begin{array}{c} \hat{\mu} \\ \hat{t}_1 \\ \hat{t}_2 \\ \hline \hat{\phi} \\ \hline \hat{s}_1 \\ \hat{s}_2 \\ \hat{s}_3 \end{array}\right] = \left[\begin{array}{c} 331 \\ 115 \\ 216 \\ \hline 4{,}33 \\ \hline 159 \\ 142 \\ 30 \end{array}\right]$$

Når restriksjonen $\hat{\mu} = 0$ blir brukt, er $\hat{\phi} = \underline{-0{,}282}$.

$\hat{k} = H\hat{\phi} = M'RM\hat{\phi} = (\frac{11}{90})\ (-0{,}282) = \underline{-0{,}035}$

Resultatene kan tyde på en "inter-treatment" seleksjon. I så fall er $\hat{t}_1 - \hat{t}_2$ ikke estimérbar, s_j er imidlertid estimérbar.

b. Et annet eksempel er gitt av Fimland (1975 b). Han forutsetter fölgende datasett:

	Besetning-år			
Far	1,1	1,2	2,1	2,2
1	2	3	1	3
2	3	0	2	0

Som en ser, har far 2 ikke avkom i år 2, fordi avkom etter denne faren kan ha gjort det dårlig i år 1. Her blir:

$$L' = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -\frac{1}{2} & -\frac{1}{2} & 0 & 0 & 0 \end{bmatrix}$$

og

$$X = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

dvs. $L'X = 0$

Med andre ord prediksjon av avlsverdiene er forventningsrette da besetning-år betraktes som "fixed".

c. Seleksjon på y:

Vi forutsetter fölgende datasett (Henderson, 1973):

Far nr.	Antall avkom	Gjennomsnitt for hver avkomsgruppe
1 (s_1)	5 og 100	y_{11} , y_{12}
2 (s_2)	4	y_{21}
3 (s_3)	2	y_{31}
4 (s_4)	5	y_{41}

Videre forutsetter vi at far nr. 1 og 2 tilhörer gruppe 1 (dvs. g_1), og far nr. 3 og 4 tilhörer gruppe 2 (dvs. g_2).

$G = I\sigma_s^2$, dvs. $G^{-1} = \dfrac{\sigma_e^2}{\sigma_s^2}$ (når $R = I$), som settes lik 5 i dette eksemplet.

En kan da sette opp BLUP-ligningene for det uselekterte tilfellet:

$$\begin{bmatrix} 109 & 0 & 105 & 4 & 0 & 0 \\ 0 & 7 & 0 & 0 & 2 & 5 \\ 105 & 0 & 110 & 0 & 0 & 0 \\ 4 & 0 & 0 & 9 & 0 & 0 \\ 0 & 2 & 0 & 0 & 7 & 0 \\ 0 & 5 & 0 & 0 & 0 & 10 \end{bmatrix} \begin{bmatrix} \hat{g}_1 \\ \hat{g}_2 \\ \hat{s}_1 \\ \hat{s}_2 \\ \hat{s}_3 \\ \hat{s}_4 \end{bmatrix} = \begin{bmatrix} y_{11} + y_{12} + y_{21} \\ y_{31} + y_{41} \\ y_{11} + y_{12} \\ y_{21} \\ y_{31} \\ y_{41} \end{bmatrix}$$

som gir fölgende lösninger:

$$\begin{bmatrix} \hat{g}_1 + \hat{s}_1 \\ \hat{g}_1 + \hat{s}_2 \\ \hat{g}_2 + \hat{s}_3 \\ \hat{g}_2 + \hat{s}_4 \end{bmatrix} = \begin{bmatrix} ,0469 & ,9386 & ,0145 & 0 & 0 \\ ,0181 & ,3610 & ,6209 & 0 & 0 \\ 0 & 0 & 0 & ,5455 & ,4545 \\ 0 & 0 & 0 & ,1818 & ,8182 \end{bmatrix} \begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{31} \\ y_{41} \end{bmatrix}$$

Det kan imidlertid tenkes at far nr. 1 ble selektert til å ha en andre omgang med avkom på grunn av prestasjonene til de förste 5 avkom (y_{11}). Dette kan uttrykkes som:

$$E \begin{bmatrix} y_{11} - y_{21} \\ y_{11} - y_{31} \\ y_{11} - y_{41} \end{bmatrix} \neq \begin{bmatrix} 0 \\ g_1 - g_2 \\ g_1 - g_2 \end{bmatrix} \quad \text{og}$$

$$L' = \begin{bmatrix} 1 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 & -1 \end{bmatrix}$$

Dette betyr at:

$$L'X = L' \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & -1 \\ 1 & -1 \end{bmatrix} \neq \text{null, og derfor gir}$$

vanlig BLUP "biased" resultater.

Videre:

$$L'VL = L' \begin{bmatrix} ,4 & ,2 & 0 & 0 & 0 \\ ,2 & ,21 & 0 & 0 & 0 \\ 0 & 0 & ,45 & 0 & 0 \\ 0 & 0 & 0 & ,7 & 0 \\ 0 & 0 & 0 & 0 & ,4 \end{bmatrix} L\sigma_e^2$$

$$= \begin{bmatrix} ,4 & ,2 & -,45 & 0 & 0 \\ ,4 & ,2 & 0 & -,7 & 0 \\ ,4 & ,2 & 0 & 0 & -,4 \end{bmatrix} L\sigma_e^2 =$$

$$= \begin{bmatrix} ,85 & ,4 & ,4 \\ ,4 & 1,1 & ,4 \\ ,4 & ,4 & ,8 \end{bmatrix} \sigma_e^2$$

Nå er:

$$\hat{b} = \begin{bmatrix} \hat{g}_1 \\ \hat{g}_2 \end{bmatrix} ; \qquad \hat{\underline{u}} = \begin{bmatrix} \hat{s}_1 \\ \hat{s}_2 \\ \hat{s}_3 \\ \hat{s}_4 \end{bmatrix} \text{ og } \hat{\underline{\phi}} = \begin{bmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \\ \hat{\phi}_3 \end{bmatrix}$$

En kan da sette opp de modifiserte BLUP-ligningene:

$$\begin{bmatrix} 109 & 0 & 105 & 4 & 0 & 0 & 0 & 1 & 1 \\ 0 & 7 & 0 & 0 & 2 & 5 & 0 & -1 & -1 \\ 105 & 0 & 110 & 0 & 0 & 0 & 0 & 0 & 0 \\ 4 & 0 & 0 & 9 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 7 & 0 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 & 0 & 10 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & ,85 & ,4 & ,4 \\ 1 & -1 & 0 & 0 & 0 & 0 & ,4 & 1,1 & ,4 \\ 1 & -1 & 0 & 0 & 0 & 0 & ,4 & ,4 & 8 \end{bmatrix} \begin{bmatrix} \hat{g}_1 \\ \hat{g}_2 \\ \hat{s}_1 \\ \hat{s}_2 \\ \hat{s}_3 \\ \hat{s}_4 \\ \hat{\phi}_1 \\ \hat{\phi}_2 \\ \hat{\phi}_3 \end{bmatrix} = \begin{bmatrix} y_{11} + y_{12} + y_{21} \\ y_{31} + y_{41} \\ y_{11} + y_{12} \\ y_{21} \\ y_{31} \\ y_{41} \\ y_{11} - y_{21} \\ y_{11} - y_{31} \\ y_{11} - y_{41} \end{bmatrix}$$

og lösningene blir:

$$\begin{bmatrix} \hat{g}_1 + \hat{s}_1 \\ \hat{g}_1 + \hat{s}_2 \\ \hat{g}_2 + \hat{s}_3 \\ \hat{g}_2 + \hat{s}_4 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ -,5555 & 1,1111 & ,4444 & 0 & 0 \\ 1,3131 & -1,7172 & ,4041 & ,5455 & ,4545 \\ ,9192 & -1,2020 & ,2828 & ,1818 & ,8182 \end{bmatrix} \begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{31} \\ y_{41} \end{bmatrix}$$

Tar en så forventningen; får vi (sml. VI 5):

$$E\left[\begin{pmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{31} \\ y_{41} \end{pmatrix} \middle| \, L'y \right] = \begin{bmatrix} g_1 \\ g_1 \\ g_1 \\ g_2 \\ g_2 \end{bmatrix} + \underbrace{\begin{bmatrix} ,4 & ,4 & ,4 \\ ,2 & ,2 & ,2 \\ -,45 & 0 & 0 \\ 0 & -,7 & 0 \\ 0 & 0 & -,4 \end{bmatrix}}_{VL} (L'VL)^{-1}L' \begin{bmatrix} y_{11} - g_1 \\ y_{12} - g_1 \\ y_{21} - g_1 \\ y_{31} - g_2 \\ y_{41} - g_2 \end{bmatrix}$$

som er nyttig informasjon når en skal finne:

$$E\left[\begin{pmatrix} \hat{g}_1 + \hat{s}_1 \\ \hat{g}_1 + \hat{s}_2 \\ \hat{g}_2 + \hat{s}_3 \\ \hat{g}_2 + \hat{s}_4 \end{pmatrix} \middle| \, L'y \right] = \begin{bmatrix} g_1 \\ g_1 \\ g_2 \\ g_2 \end{bmatrix} + \underbrace{\begin{bmatrix} ,2 & ,2 & ,2 \\ -,2 & 0 & 0 \\ 0 & -,2 & 0 \\ 0 & 0 & -,2 \end{bmatrix}}_{GZ'L} (L'VL)^{-1}L' \begin{bmatrix} y_{11} - g_1 \\ y_{12} - g_1 \\ y_{21} - g_1 \\ y_{31} - g_2 \\ y_{41} - g_2 \end{bmatrix}$$

Det er lett å bevise at:

$$E\left[\begin{array}{c|c} g_1 + s_1 & \\ g_1 + s_2 & \\ g_2 + s_3 & L'y \\ g_2 + s_4 & \end{array}\right] \quad \text{har samme forventning.}$$

d. <u>Seleksjon på u</u>:

La oss forutsette fölgende modell og datasett:

$y_{ij} = \mu + s_i + e_{ij}$, hvor

$$\text{Var.-Kov.} \begin{pmatrix} s \\ e \end{pmatrix} = \begin{bmatrix} \frac{1}{10} I & 0 \\ 0 & I \end{bmatrix} \sigma_e^2 .$$

Far	Antall avkom (n_i)
1	10
2	8
3	25
4	5

I det <u>uselekterte</u> tilfellet er BLUP av $\mu + s_i$ lik:

$$\begin{bmatrix} \hat{\mu} + \hat{s}_1 \\ \cdot \\ \cdot \\ \cdot \\ \hat{\mu} + \hat{s}_4 \end{bmatrix} ,$$

hvor $\hat{\mu}$, $\hat{s}_i$ er lösningene til:

$$\begin{bmatrix} 48 & 10 & 8 & 25 & 5 \\ 10 & 20^* & 0 & 0 & 0 \\ 8 & 0 & 18 & 0 & 0 \\ 25 & 0 & 0 & 35 & 0 \\ 5 & 0 & 0 & 0 & 15 \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{s}_1 \\ \hat{s}_2 \\ \hat{s}_3 \\ \hat{s}_4 \end{bmatrix} = \begin{bmatrix} y_{..} \\ y_{1.} \\ y_{2.} \\ y_{3.} \\ y_{4.} \end{bmatrix}$$

$^*10 + G^{-1}$

Fölgelig:

$$\begin{bmatrix} \hat{\mu} + \hat{s}_1 \\ \hat{\mu} + \hat{s}_2 \\ \hat{\mu} + \hat{s}_3 \\ \hat{\mu} + \hat{s}_4 \end{bmatrix} = \begin{bmatrix} ,6255 & ,1115 & ,1793 & ,0837 \\ ,1394 & ,5684 & ,1992 & ,0930 \\ ,0717 & ,0638 & ,8167 & ,0478 \\ ,1673 & ,1488 & ,2390 & ,4449 \end{bmatrix} \begin{bmatrix} \bar{y}_{1.} \\ \bar{y}_{2.} \\ \bar{y}_{3.} \\ \bar{y}_{4.} \end{bmatrix}$$

Nå kan det tenkes at $E[s_1 + s_2 - (s_3 + s_4)] \neq 0$ på grunn av seleksjon. Da blir:

$$L' = [1 \quad 1 \quad -1 \quad -1] \qquad \text{og}$$

$$L'\underline{u} = [s_1 + s_2 - s_3 - s_4]$$

BLUP-ligningene nå:

$$\begin{bmatrix} 48 & 10 & 8 & 25 & 5 & 0 \\ 10 & 20 & 0 & 0 & 0 & -1 \\ 8 & 0 & 18 & 0 & 0 & -1 \\ 25 & 0 & 0 & 35 & 0 & 1 \\ 5 & 0 & 0 & 0 & 15 & 1 \\ 0 & -1 & -1 & 1 & 1 & ,4 \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{s}_1 \\ \hat{s}_2 \\ \hat{s}_3 \\ \hat{s}_4 \\ \hat{\phi} \end{bmatrix} = \begin{bmatrix} y_{..} \\ y_{1.} \\ y_{2.} \\ y_{3.} \\ y_{4.} \\ 0 \end{bmatrix}$$

og lösningene blir:

$$\begin{bmatrix} \hat{\mu} + \hat{s}_1 \\ \hat{\mu} + \hat{s}_2 \\ \hat{\mu} + \hat{s}_3 \\ \hat{\mu} + \hat{s}_4 \end{bmatrix} = \begin{bmatrix} ,7647 & ,2353 & 0 & 0 \\ ,2941 & ,7059 & 0 & 0 \\ 0 & 0 & ,9091 & ,0909 \\ 0 & 0 & ,4545 & ,5455 \end{bmatrix} \begin{bmatrix} \bar{y}_{1.} \\ \bar{y}_{2.} \\ \bar{y}_{3.} \\ \bar{y}_{4.} \end{bmatrix}$$

En bör merke seg her at $\hat{\mu} + \hat{s}_i$ gir ulike verdier i det uselekterte og i det selekterte tilfellet.

Et annet sett å oppnå "unbiased" resultater på er å gruppere far 1 og 2 i én gruppe (g_1) og far 3 og 4 i en annen gruppe (g_2). Da blir BLUP-ligningene:

$$\begin{bmatrix} 18 & 0 & 10 & 8 & 0 & 0 \\ 0 & 30 & 0 & 0 & 25 & 5 \\ 10 & 0 & 20 & 0 & 0 & 0 \\ 8 & 0 & 0 & 18 & 0 & 0 \\ 0 & 25 & 0 & 0 & 35 & 0 \\ 0 & 5 & 0 & 0 & 0 & 15 \end{bmatrix} \begin{bmatrix} \hat{g}_1 \\ \hat{g}_2 \\ \hat{s}_1 \\ \hat{s}_2 \\ \hat{s}_3 \\ \hat{s}_4 \end{bmatrix} = \begin{bmatrix} y_{1.} + y_{2.} \\ y_{3.} + y_{4.} \\ y_{1.} \\ y_{2.} \\ y_{3.} \\ y_{4.} \end{bmatrix}$$

$\hat{g} + \hat{s}$ blir da lik $\hat{\mu} + \hat{s}$ i seleksjonsmodellen.

e. Seleksjon på y:

Fölgende datasett forutsettes (Henderson, 1975 e):

År-sesong (s)	Ku 1	2	4	5	6	Sum
1	1068	1959	730	...	...	3757
2	1355	1578	...	...	...	2933
3	...	1204	...	1603	1810	4617
Sum	2423	4741	730	1603	1810	11307

Modellen som ligger til grunn for den enkelte registrering kan sies å väre:

$$\underline{y} = X\underline{b} + Z\underline{a} + Z\underline{p} + \underline{e} \quad , \text{ hvor}$$

$\underline{y}$ = vektor av registreringer (observasjoner)

$\underline{b}$ = vektor av år-sesong effekter

$\underline{a}$ = vektor av additive genetiske verdier

$\underline{p}$ = vektor av ikke additive genetiske verdier samt effekt av permanent miljö

$\underline{e}$ = vektor av tilfeldige miljöeffekter, osv.

Videre er: $\sigma_a^2 = Ah^2\sigma_y^2$ hvor A er matrisen for det additive genetiske slektskapet mellom individene.

$$\sigma_p^2 = I(r - h^2)\sigma_y^2 \quad ; \quad \sigma^2 = I(1 - r)\sigma_y^2.$$

Dette betyr at: $\sigma_y^2 = [\, ZAh^2Z' + ZI(r - h^2)Z' + I(1 - r)]\sigma_y^2$,
og $\underline{a}$, $\underline{p}$ og $\underline{e}$ forutsettes å väre ukorrelerte.

Vi kan da sette opp det generelle uttrykket for "mixed model"-ligningene:

$$\begin{bmatrix} X'X & X'Z & X'Z \\ Z'X & Z'Z + \frac{(1-r)}{h^2}A^{-1} & Z'Z \\ Z'X & Z'Z & Z'Z + \frac{(1-r)}{(r-h^2)}I \end{bmatrix} \begin{bmatrix} \hat{\underline{b}} \\ \hat{\underline{a}} \\ \hat{\underline{p}} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ Z'y \end{bmatrix}$$

For vårt eksempel er:

$$\underline{y}' = [y_{11} \quad y_{12} \ \ldots \ y_{36}],$$

$$X = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\underline{b} = \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix}$$

$$Z = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Legg merke til at tredje kolumnen i Z er 0 fordi individ 3 ikke har noen observasjon.

Når det gjelder relasjoner mellom individene forutsettes fölgende (pil går fra foreldre til avkom):

$1 \longrightarrow 5$

$2 \searrow$
$\quad 6$
$3 \nearrow$

4 , som gir

$$A^{-1} = \frac{1}{6}\begin{bmatrix} 8 & 0 & 0 & 0 & -4 & 0 \\ 0 & 9 & 3 & 0 & 0 & -6 \\ 0 & 3 & 9 & 0 & 0 & -6 \\ 0 & 0 & 0 & 6 & 0 & 0 \\ -4 & 0 & 0 & 0 & 8 & 0 \\ 0 & -6 & -6 & 0 & 0 & 12 \end{bmatrix}$$

En får da fölgende BLUP-ligninger (NB $\hat{p}_3$ ligningen er tatt bort) når $h^2 = 0{,}25$, $r = 0{,}4$, $\frac{1-r}{h^2} = 2{,}4$ og $\frac{1-r}{r-h^2} = 4$:

$$\left[\begin{array}{ccc|cccccc|ccccc}
3{,}0 & {,}0 & {,}0 & 1{,}0 & 1{,}0 & {,}0 & 1{,}0 & {,}0 & {,}0 & 1{,}0 & 1{,}0 & 1{,}0 & {,}0 & {,}0 \\
{,}0 & 2{,}0 & {,}0 & 1{,}0 & 1{,}0 & {,}0 & {,}0 & {,}0 & {,}0 & 1{,}0 & 1{,}0 & {,}0 & {,}0 & {,}0 \\
{,}0 & {,}0 & 3{,}0 & {,}0 & 1{,}0 & {,}0 & {,}0 & 1{,}0 & 1{,}0 & {,}0 & 1{,}0 & {,}0 & 1{,}0 & 1{,}0 \\
\hline
1{,}0 & 1{,}0 & {,}0 & 5{,}2 & {,}0 & {,}0 & {,}0 & -1{,}6 & {,}0 & 2{,}0 & {,}0 & {,}0 & {,}0 & {,}0 \\
1{,}0 & 1{,}0 & 1{,}0 & {,}0 & 6{,}6 & 1{,}2 & {,}0 & {,}0 & -2{,}4 & {,}0 & 3{,}0 & {,}0 & {,}0 & {,}0 \\
{,}0 & {,}0 & {,}0 & {,}0 & 1{,}2 & 3{,}6 & {,}0 & {,}0 & -2{,}4 & {,}0 & {,}0 & {,}0 & {,}0 & {,}0 \\
1{,}0 & {,}0 & {,}0 & {,}0 & {,}0 & {,}0 & 3{,}4 & {,}0 & {,}0 & {,}0 & {,}0 & 1{,}0 & {,}0 & {,}0 \\
{,}0 & {,}0 & 1{,}0 & -1{,}6 & {,}0 & {,}0 & {,}0 & 4{,}2 & {,}0 & {,}0 & {,}0 & {,}0 & 1{,}0 & {,}0 \\
{,}0 & {,}0 & 1{,}0 & {,}0 & -2{,}4 & -2{,}4 & {,}0 & {,}0 & 5{,}8 & {,}0 & {,}0 & {,}0 & {,}0 & 1{,}0 \\
\hline
1{,}0 & 1{,}0 & {,}0 & 2{,}0 & {,}0 & {,}0 & {,}0 & {,}0 & {,}0 & 6{,}0 & {,}0 & {,}0 & {,}0 & {,}0 \\
1{,}0 & 1{,}0 & 1{,}0 & {,}0 & 3{,}0 & {,}0 & {,}0 & {,}0 & {,}0 & {,}0 & 7{,}0 & {,}0 & {,}0 & {,}0 \\
1{,}0 & {,}0 & {,}0 & {,}0 & {,}0 & {,}0 & 1{,}0 & {,}0 & {,}0 & {,}0 & {,}0 & 5{,}0 & {,}0 & {,}0 \\
{,}0 & {,}0 & 1{,}0 & {,}0 & {,}0 & {,}0 & {,}0 & 1{,}0 & {,}0 & {,}0 & {,}0 & {,}0 & 5{,}0 & {,}0 \\
{,}0 & {,}0 & 1{,}0 & {,}0 & {,}0 & {,}0 & {,}0 & {,}0 & 1{,}0 & {,}0 & {,}0 & {,}0 & {,}0 & 5{,}0
\end{array}\right]
\begin{bmatrix} \hat{s}_1 \\ \hat{s}_2 \\ \hat{s}_3 \\ \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \\ \hat{a}_5 \\ \hat{a}_6 \\ \hat{p}_1 \\ \hat{p}_2 \\ \hat{p}_4 \\ \hat{p}_5 \\ \hat{p}_6 \end{bmatrix}
=
\begin{bmatrix} 3757 \\ 2933 \\ 4617 \\ 2423 \\ 4741 \\ 0 \\ 730 \\ 1603 \\ 1810 \\ 2423 \\ 4741 \\ 730 \\ 1603 \\ 1810 \end{bmatrix}$$

og lösningene:

$$\hat{s} = [1306{,}5 \quad 1432{,}4 \quad 1412{,}0];$$

$$\hat{a} = [-38{,}4 \quad 108{,}7 \quad 45{,}8 \quad -144{,}1 \quad 22{,}8 \quad 123{,}1];$$

$$\hat{p} = [-39{,}9 \quad 37{,}7 \quad -86{,}5 \quad 33{,}6 \quad 55{,}0].$$

Nå kan det tenkes at ku nr. 4 ble utrangert fordi hennes förste registrering (produksjon) var lågere enn gjennomsnittet for ku nr. 1 og 2. Videre kan det tenkes at ku nr. 2 ble holdt for et tredje produksjonsår fordi de to förste var större enn for ku nr. 1. Dette betyr at:

$$E\begin{bmatrix} y_{11} + y_{12} - 2y_{14} \\ -y_{11} + y_{12} - y_{21} + y_{22} \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = L'\underline{y} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

og

$$L' = \begin{bmatrix} 1 & 1 & -2 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & -1 & 1 & 0 & 0 & 0 \end{bmatrix}$$

Spörsmålet blir da: Er $\hat{\underline{s}}$, $\hat{\underline{a}}$ og $\hat{\underline{p}}$ "unbiased"?

Det enkleste settet å besvare dette spörsmålet på er å uttrykke lösningene for henholdsvis $\hat{\underline{s}}$, $\hat{\underline{a}}$ og $\hat{\underline{p}}$ i form av $\underline{y}$ og så se om $E(\hat{\underline{s}}|L'\underline{y}) = E(\underline{s}|L'\underline{y})$; $E(\hat{\underline{a}}|L'\underline{y}) = E(\underline{a}|L'\underline{y})$ og $E(\hat{\underline{p}}|L'\underline{y}) = E(\underline{p}|L'\underline{y})$. För en gjör dette er det av stor hjelp å ha $E(\underline{y}|L'\underline{y})$ tilgjengelig (sml. VI 5).

Ved å fölge opplegget her vil en for dette eksemplet finne at seleksjonen ikke förer til "bias" i $\hat{\underline{s}}$, $\hat{\underline{a}}$ og $\hat{\underline{p}}$, med andre ord har en ved bruk av A^{-1} og BLUP klart å eliminere "bias" på grunn av seleksjon av bl.a. mödre. En annen fremgangsmåte ville det ha värt å gruppere kyrne i generasjoner og så anvende BLUP.

Flere eksempler er gitt av Henderson (1975 a) hvor det blant annet er illustrert betydningen av å kjenne til om effektene er "fixed" eller "random" når en tar forventningen.

VII. BEREGNING AV VARIANSKOMPONENTER FRA "MIXED-MODEL"-LIGNINGENE - EN ITERATIV METODE

Som tidligere påpekt forutsetter BLUP-metoden egentlig at varianser og kovarianser er kjente. Om varianser og kovarianser er ukjente, kan en fremdeles bruke "mixed-model"-ligningene og samtidig beregne varianskomponentene ved bruk av iterasjon. Dette er beskrevet mer utförlig av bl.a. Schaeffer (1974) og Fimland (1975 a). La modellen väre:

$$\underline{y} = X\underline{b} + Z\underline{u} + \underline{e}$$

hvor

X er en kjent, $n \times p$ matrise med rank $= r \leq \min(n, p)$;

Z er en kjent, $n \times q$ matrise

Videre,

$$E \begin{bmatrix} \underline{y} \\ \underline{u} \\ \underline{e} \end{bmatrix} = \begin{bmatrix} X\underline{b} \\ \underline{0} \\ \underline{0} \end{bmatrix} \quad \text{og} \quad \text{Var-Kovar.} \begin{bmatrix} \underline{y} \\ \underline{u} \\ \underline{e} \end{bmatrix} = \begin{bmatrix} V & ZG & R \\ GZ' & G & 0 \\ R & 0 & R \end{bmatrix}$$

For å presisere närmere er:

$$\underline{u}' = [u_1' \; \dots \; u_s'] \quad \text{og}$$

$$Z = [Z_1 \; \dots \; Z_s] \quad ,$$

$$R = I\sigma_o^2 \quad \text{og} \quad G = \Sigma^+ \; I\sigma_i^2 .$$

Videre la $\tilde{\sigma}_o^2$ og $\tilde{\sigma}_i^2$ $(i = 1 \dots s)$ väre de förste verdiene på de ukjente varianskomponentene.

Vi kan nå sette opp "mixed-model"-ligningene:

$$\begin{bmatrix} X'X & X'Z_1 & \dots & X'Z_s \\ Z_1'X & Z_1'Z_1 + I\tilde{\sigma}_o^2/\tilde{\sigma}_1^2 & \dots & Z_1'Z_s \\ \vdots & & & \vdots \\ X'Z_s & Z_s'Z_1 & \dots & Z_s'Z_s + I\tilde{\sigma}_o^2/\tilde{\sigma}_s^2 \end{bmatrix} \begin{bmatrix} \hat{\underline{b}} \\ \hat{u}_1 \\ \vdots \\ \hat{u}_s \end{bmatrix} = \begin{bmatrix} X'y \\ Z_1'y \\ \vdots \\ Z_s'y \end{bmatrix}$$

La den inverse matrisen väre:

$$\begin{bmatrix} T_{oo} & T_{o1} \cdots & T_{os} \\ T'_{o1} & T_{11} \cdots & T_{1s} \\ \vdots & \vdots & \vdots \\ T'_{os} & T'_{1s} \cdots & T_{ss} \end{bmatrix}$$

Dette innebärer at:

$$t_o = y'y - \hat{\beta}'X'y - \hat{u}'Z'y - \sum_{i=1}^{s} \hat{u}_i' \hat{u}_i \, \tilde{\sigma}_o^2/\tilde{\sigma}_i^2 .$$

$$t_i = \hat{u}_i' \hat{u}_i \, (\tilde{\sigma}_o^2/\tilde{\sigma}_i^2)^2 \text{ , for } i = 1, \ldots, s$$

Videre at $\tilde{\sigma}^2 = P^{-1}t$ hvor elementene av P er:

$$P_{oo} = n - r - q + \sum_{i=1}^{s} \sum_{j=1}^{s} \operatorname{tr} T_{ij} T_{ij}' \frac{\tilde{\sigma}_o^2 \tilde{\sigma}_o^2}{\tilde{\sigma}_i^2 \tilde{\sigma}_j^2} .$$

$$P_{oi} = \left(\frac{\tilde{\sigma}_o^2}{\tilde{\sigma}_i^2}\right)^2 [\operatorname{tr} T_{ii} - \sum_{j=1}^{s} \operatorname{tr} T_{ij} T_{ij}' \frac{\tilde{\sigma}_o^2}{\tilde{\sigma}_j^2}] \text{, for } i = 1, \ldots, s$$

$$P_{ii} = \left(\frac{\tilde{\sigma}_o^2}{\tilde{\sigma}_i^2}\right)^2 \{q_i - 2 \operatorname{tr} T_{ii} \frac{\tilde{\sigma}_o^2}{\tilde{\sigma}_i^2} + \operatorname{tr} (T_{ii})^2 \left(\frac{\tilde{\sigma}_o^2}{\tilde{\sigma}_i^2}\right)^2 \} \text{, for } i = 1, \ldots, s$$

$$P_{ij} = \left(\frac{\tilde{\sigma}_o^2}{\tilde{\sigma}_i^2}\right)^2 \operatorname{tr} T_{ij} T_{ij}' \left(\frac{\tilde{\sigma}_o^2}{\tilde{\sigma}_j^2}\right)^2 \text{, for } i,j = 1, \ldots, s \text{ og } i \neq j$$

Ved å sette de nye estimatene for $\tilde{\sigma}^2$ inn på diagonalene i "mixed-model"-ligningene kan en få et bedre estimat på σ_o^2 og σ_i^2. Vanligvis vil 10 iterasjoner väre tilstrekkelig. $\tilde{\sigma}_o^2$ og $\tilde{\sigma}_i^2$ har da nesten de samme "properties" som MINQUE ("Minimum Norm Quadratic Unbiased Estimators") av varianskomponentene, og lösningen for $\underline{u}$ er tilnärmet BLUP. Et problem med iterasjon er at en kan få negative varianskomponenter (Fimland, 1975 a).

Om MINQUE skal bli BQUE ("Best Quadratic Unbiased Estimator"), må en forutsette normalitet. I så fall er BLUP lik BP ("best prediction").

VIII. "REGRESSED LEAST-SQUARES MEANS"

Denne metoden ble foreslått av Cunningham (1965) for rangering av potensielle avlsdyr. Etter som metoden blir relativt mye brukt i dag, skal vi gå noe närmere inn på denne.

La oss gå ut fra fölgende modell:

$$Y_{ijk} = \mu + s_i + h_j + e_{ijk}$$

hvor

μ = konstant

s_i = effekt av i^{te} hanndyr ("fixed" effekt)

h_j = effekt av j^{te} besetning (el. besetning-år-sesong kombinasjon) $[0, \sigma_h^2]$. Altså er besetningseffekten betraktet som "random".

e_{ijk} = tilfeldig effekt på "record" til k^{te} datter etter i^{te} far i j^{te} besetning

Denne modellen kan skrives om på fölgende sett (Fimland, 1976 b):

$$\underline{y} = K\underline{s}^* + X_1'\underline{h} + \underline{e}$$

hvor

$$\underline{s}^* = \underline{\mu} + \underline{s}$$

De tradisjonelle normalligningene blir da:

$$\begin{bmatrix} K'K & K'X_1 \\ X_1'K & X_1'X+I\delta \end{bmatrix} \begin{bmatrix} \hat{\underline{s}}^* \\ \hat{\underline{h}} \end{bmatrix} = \begin{bmatrix} K'\underline{y} \\ X_1'\underline{y} \end{bmatrix} \qquad \text{og}$$

$$\delta = \frac{\sigma_e^2}{\sigma_h^2} \quad \text{(forutsettes kjent)}$$

Videre la den inverse koeffisientmatrisen väre:

$$\begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} = \begin{bmatrix} K'K & K'X_1 \\ X_1'K & X_1'X_1+I\delta \end{bmatrix}^{-1}$$

Med de definisjonene som er brukt, er koeffisientmatrisen av full rank.

Som vist av Cunningham (1965) kan en beregne avlsverdien for det enkelte hanndyr på fölgende måte:

$$\hat{\hat{\underline{s}}}^* = \hat{\mu} + d\ (\hat{\underline{s}}^* - \hat{\mu})$$

hvor

$$d = \frac{2\ \sigma_s^2}{\sigma_s^2 + C^{ii}\ \sigma_e^2}$$

Gir denne metoden et forventningsrett estimat av s^* ? Den enkleste måten å undersöke dette på er å uttrykke estimatene som en lineär funksjon av elementene i modellen. Således:

$$\begin{aligned}\hat{\underline{s}}^* &= C_{11}\ K'\underline{y} + C_{12}\ X_1'\underline{y} = (C_{11}\ K' + C_{12}\ X_1')\underline{y} = \\ &(C_{11}\ K' + C_{12}\ X_1')(K\underline{s}^* + X_1\underline{h} + \underline{e}) = \\ &(C_{11}\ K'K + C_{12}\ X_1'K)\underline{s}^* + (C_{11}\ K'X_1 + C_{12}\ X_1'X_1)\underline{h} + \\ &(C_{11}\ K' + C_{12}\ X_1')\underline{e} = \underline{s}^* - C_{12}\ I\delta\underline{h} + (C_{11}\ K' + C_{12}\ X_1')\underline{e}\end{aligned}$$

fordi,

$$C_{11}\ K'K + C_{12}\ X_1'K = I$$

$$C_{11}\ K'X_1 + C_{12}\ (X_1'X_1 + I\delta) = 0$$

$$C_{11}\ K'X_1 + C_{12}\ X_1'X_1 = -C_{12}\ I\delta$$

La oss videre forutsette at $E(\underline{e}) = 0$. Dette betyr at et forventningsrett estimat av $\underline{s}^*$ vil en oppnå med denne metoden i fölgende to tilfeller:

° balansert "design"

°° helt tilfeldig fordeling av fedre på besetninger

dvs. da blir C_{12} Iδh lik 0.

I storfeavlen i Sverige hvor eliteparingene har en tendens til å forekomme oftere i de högstytende besetningene kan ikke denne metoden sies å väre noe videre hövelig.

En utförligere diskusjon om denne metoden og andre metoder for rangering av potensielle avlsdyr er gitt av bl.a. Eimland (1976 b). Videre diskuterer Cunningham (1965) metoden når både $\underline{s}^*$ og $\underline{h}$ betraktes som "fixed".

IX. NOEN BETRAKTNINGER OMKRING OPPBYGGING AV MODELLER

Et spörsmål som ofte reises, er hvilke feil det innebärer å bruke feilaktige modeller. Dette spörsmålet er selvsagt komplisert og omfattende, og en skal ikke her gi noen utförlig diskusjon. Imidlertid kan vi forsöke å summere opp noen fakta med hensyn til dette problemet. Bl.a. har Hendersen (1975 b) gitt en bra oversikt her. Interesserte henvises derfor til denne publikasjonen for en mer utförlig redegjörelse.

∘ Konsekvenser av å inkludere "fixed"-effekter av ingen betydning:

La oss anta at den riktige modellen er: $\underline{y} = X_1\underline{b}_1 + Z\underline{u} + \underline{e}$, mens vi bruker modellen:

$$\underline{y} = X_1\underline{b}_1 + X_2\underline{b}_2 + Z\underline{u} + \underline{e}$$

Bruk av den feilaktige modellen forårsaker ikke "bias" i estimatorene og prediktorene. Derimot vil visse funksjoner som er estimérbare når en bruker förstnevnte modell, ikke väre estimérbare når den feilaktige modellen anvendes, videre vil sampling-variansen bli större.

∘∘ Konsekvenser av å ignorere "fixed"-effekter av betydning:

Her blir både estimatorer og prediktorer "biased", men sampling-variansen blir mindre ved bruk av den feilaktige modellen.

∘∘∘ Konsekvenser av å ignorere "random"-effekter av betydning:

Her forutsetter en at den riktige modellen er:

$$\underline{y} = X\underline{b} + Z_1\underline{u}_1 + Z_2\underline{u}_2 + \underline{e}$$, men en bruker:

$$\underline{y} = X\underline{b} + Z_1\underline{u}_1 + \underline{e}$$

Estimatorene og prediktorene blir "unbiased", men sampling-variansen blir större ved bruk av den feilaktige modellen.

.... <u>Konsekvensen av å bruke feilaktig forhold mellom varians-komponentene (G)</u>:

Da blir variansen på $k'\hat{\underline{b}} + m'\hat{\underline{u}}$ större enn om riktig G anvendes.

..... <u>Konsekvenser av å behandle "random"-elementene som "fixed"-effekter</u>:

Dette medförer ingen "bias", men variansen på "prediction errors" blir större enn ved bruk av riktig modell.

...... <u>Konsekvensen av å ikke bruke alle data som er tilgjengelige</u>:

Her kan en tenke seg at det er enklere å bruke BLUP om en f.eks. rangerer oksene på grunnlag av mjölkeavkastningen i förste laktasjonen dvs. en ignorerer mjölkeavkastningen i senere laktasjoner. Ved å gjöre dette så aksepteres en högere varians på forskjellen mellom estimatene av okseeffektene.

Litteratur

Cunningham, E.P. 1965. The evaluation of sires from progeny test data. Anim. Prod. 7: 221-231.

Fimland, E.A. 1972. Analyse av en generell lineär modell. Skrevet for nordisk lisensiatkurs i populasjonsgenetikk, Ås-NLH: 1-56.

Fimland, E.A. 1975 a. Estimation of sire's breeding value. III. An iterative procedure for estimation of linear parameters and heteroscedastic variances simultaneously. Z. Tierzücht. ZüchtBiol. 92: 10-16.

Fimland, E.A. 1975 b. Estimation of sire's breeding value. IV. Maximum likelihood derivation of the predictor and the estimator of a general mixed model given some type of selection of a correlated random variable. Z. Tierzücht. ZüchtBiol. 92: 176-187.

Fimland, E.A. 1976 a. Estimation of a sire's breeding value. V. Derivations of variance - covariance matrices of a solution of normal equations of a mixed linear model subject to selection bias. Z. Tierzücht. ZüchtBiol. 93: 3-13.

Fimland, E.A. 1976 b. Estimation of a sire's breeding value. VI. Different linear prediction methods discussed in the context of the breeding structure of the population. Z. Tierzücht. ZüchtBiol. 93: 14- 30.

Graybill, F.A. 1961. Introduction to Linear Statistical Models. Volume I. McGraw-Hill Book Company, Inc., London.

Harvey, W.R. 1960. Least squares analysis of data with unequal subclass numbers. U.S.D.A. ARS 20-8.

Henderson, C.R. 1963. Selection index and expected genetic advance. In: Statistical Genetics and Plant Breeding. Hanson, W.D. and Robinson, H.F. (Eds.): 141-163.

Henderson, C.R. 1968. Lecture notes for the course "An. Sci. 520". Dept. of Animal Science, Cornell University, Ithaca, New York.

Henderson, C.R. 1973. Sire evaluation and genetic trends. Proceedings of the Animal Breeding and Genetics Symposium in honor of Dr. Jay L. Lush, held July 29, 1972, at Virginia Polytechnic Institute and State University, Blacksburg, Virginia.: 10-41.

Henderson, C.R. 1975 a. Best linear unbiased estimation and prediction under a selection model. Biometrics 31: 423-447.

Henderson, C.R. 1975 b. Comparison of alternative sire evaluation methods. J. Anim. Sci. 41: 760-770.

Henderson, C.R. 1975 c. Rapid method for computing the inverse of a relationship matrix. J. Dairy Sci. 58: 1727-1730.

Henderson, C.R. 1975 d. Use of relationships among sires to increase accuracy of sire evaluation. J. Dairy Sci. 58: 1731-1738.

Henderson, C.R. 1975 e. Use of all relatives in intraherd prediction of breeding values and producing abilities. J. Dairy Sci. 58: 1910-1916.

Henderson, C.R. 1975 f. Inverse of a matrix of relationships due to sires and maternal grandsires. J. Dairy Sci. 58: 1917-1921.

Henderson, C.R., Kempthorne, O., Searle, S.R. & von Krosigk, C.M. 1959. The estimation of environmental and genetic trends from records subject to culling. Biometrics 15: 192-218.

Rönningen, K. 1971. Estimering av bias på grunn av seleksjon. Institutt for husdyravl, Norges Landbrugshögskole. Stencil: 1-19.

Schaeffer, L.R. 1974. The use of mixed models in genetic analysis. Dept. of Animal and Poultry Science, University of Guelph, Guelph, Ontario. Canada.

Schaeffer, L.R. 1975. Dairy sire evaluation for milk and fat production. Dept. of Animal and Poultry Science, University of Guelph, Guelph. Ontario. Canada.

Searle, S.R. 1967. Matrix algebra for the biological sciences (including applications in statistics). John Wiley & Sons, Inc.: 1-296.

Searle, S.R. 1971. Linear models. John Wiley & Sons, Inc.: 1-532.

Smith, C. 1962. Estimation of genetic change in farm livestock using field records. Anim. Prod. 4: 239-251.

Van Vleck, L.D. 1975. BLUP-examples. An International dr. course in animal breeding at the Dept. of Animal Breeding, The University of Agriculture, Forestry and Veterinary Medicine, August 1975. Uppsala, Sweden.