

Noter til Specialkursus i videregående statistik

Poul Thyregod

IMM, februar 2005

Indhold

Forord	6
1 Momenter og flerdimensionale stokastiske variable	7
1.0 Indledning	7
1.1 Notation	7
1.2 Momenter	8
1.3 Momenter for flerdimensionale variable	10
1.4 Fordelinger afledt af normalfordelingen	14
1.4.1 The log-normal distribution model	14
1.4.2 The χ^2 -distribution model	15
1.4.3 Relation to the normal distribution model	16
1.4.4 The t-distribution model	18
1.4.5 The F-distribution model	21
1.4.6 The Gamma-function	22
1.4.7 Oversigt over en række sædvanlige fordelinger	23
2 Likelihoodfunktion	24
2.0 Introduktion og oversigt	24
2.1 Likelihoodfunktion, Scorefunktion og Informationsmatrix	25
2.2 Eksempler	30
3 Lidt om den generelle lineære model (normalfordelingsteori)	38
3.0 Introduktion og oversigt	38
3.1 Statistisk model, den flerdimensionale normalfordelingsmodel .	39
3.2 Den generelle lineære model	42
3.3 Estimation i en GLM	45

3.4	Fordeling af $\hat{\beta}$, fittede værdier og residualer, estimation af σ^2	48
3.4.1	Fordeling af $\hat{\beta}$, fittede værdier og residualer	48
3.4.2	Kvadratiske former, spaltningssætningen	49
3.4.3	Fordeling af residualkvadratsummen, estimation af σ^2	50
3.4.4	Test af enkelte parametre	51
3.5	Eksempel, simpel lineær regressionsmodel	51
3.6	Reduktion af en generel lineær model	54
3.6.1	Introduktion	54
3.6.2	Kvotienttest for modelreduktion	55
3.6.3	Testet udtrykt i en parametriseret model	60
3.6.4	Modelreduktion ved partielle tests	62
3.6.5	Successiv testning, type I opspaltning	64
3.7	Repræsentation af modeller, modelformler og analytiske udtryk	66
3.7.1	Intercept led	68
3.7.2	Kontinuerte kovariable	69
3.7.3	Kvalitative kovariable, faktorvariable	70
3.7.4	Modelformler	74
3.7.5	Inklusionsdiagrammer for hypotese kæder, Type I og Type III opspaltning	77
3.8	Eksempel, Blodgennemstrømningsindeks hos malere	81
3.9	Vurdering af modeltilpasning, diagnostiske størrelser	84
3.9.1	Residualer, standardisering og studentisering	84
3.9.2	Kontrol af enkeltobservationer, leverage	86
3.9.3	Kontrol af enkeltobservationers overensstemmelse, residual	89
3.9.4	Kontrol af enkeltobservationers indflydelse (influens)	90
3.10	Ortogonalitet, leverage, Lille eksempel og øvelse	93
3.11	Indskud om Yates' algoritme:	93
4	Introduktion til generaliserede lineære modeller	98
4.0	Oversigt	98
4.1	Baggrund og hovedideer	100
4.2	Motiverende eksempel	101
4.3	Naturlige eksponentielle familier, eksponentielle dispersionsfamilier	104
4.3.1	Naturlige eksponentielle familier	104

4.3.2	Ekspontielle dispersionsfamilier	106
4.3.3	Middelværdiafbildning, kanonisk link, variansfunktion, og enhedsdevians	109
4.4	Model, likelihood og scorefunktion	116
4.4.1	Model for et observationssæt	116
4.4.2	Likelihoodfunktion, scorefunktion og informationsmatrix	117
4.5	Generaliserede lineære modeller	119
4.5.1	Generaliseret lineær model, dimension og modelmatrix	119
4.5.2	Komponenter i en generaliseret lineær model	121
4.5.3	Fuld model	122
4.5.4	Lokal design matrix	123
4.6	Maksimum likelihood estimation	125
4.6.1	Estimation	125
4.6.2	Fordeling af maksimum-likelihood estimat	126
4.6.3	Eksempel på endimensional regression med kanonisk link	128
4.6.4	Fejlophobningsloven, iterativt genvægtet mindste kva- draters metode	133
4.7	Fordeling af fittede værdier og residualer	134
4.7.1	Fittede værdier og residualer	134
4.7.2	Fordeling af fittede værdier og residualer	136
4.7.3	Andre former for residualer	138
4.8	Test for modeltilpasning	141
4.9	Test af enkelte parametre	144
4.10	Test af delhypotese	146
4.11	Successiv testning	151
4.11.1	Eksempel, tofaktor model med binomialt respons . . .	152
4.12	Modeller med overdispersion	161
4.12.1	Quasi-devians og quasi log-likelihood	162
4.13	Værktøjer i SAS og S-plus	163
4.13.1	Generaliserede lineære modeller i SAS Insight	163
4.13.2	Andre SAS-procedurer	165
4.13.3	Generaliserede lineære modeller i S-plus	166
4.14	Eksempler på “regressionsmodeller”	166
4.14.1	Logistisk link ved binomial respons, odds	166
4.14.2	Andre linkfunktioner for binært respons	168
4.14.3	Indskud om sensitivitet, specificitet af klassifikation, ROC-kurve	173

4.14.4	Ordnet kategorisk respons	174
4.14.5	Poisson regression, brug af OFFSET	182
4.14.6	Empiriske varianser for normalfordelte observationer	183
4.14.7	Bartlett's test	188
4.15	Tofaktormodeller, vekselvirkning	188
4.15.1	Grafisk kontrol, profilplot	189
4.15.2	Tofaktormodel for Poissonfordelte data	191
4.15.3	Vekselvirkning og valg af linkfunktion	194
4.16	Linkfunktioner	202
4.17	Lidt mere om eksponentielle familier	203
5	Normalfordelingsmodeller med hierarkisk variation	213
5.1	Betingede, simultane og marginale fordelinger	214
5.1.1	Lidt om betingede og marginale middelværdier og varianser	215
5.2	Envejsanalyse, tilfældig model	216
5.2.1	Eksempel	216
5.3	Model	217
5.3.1	Marginale og simultane fordelinger	218
5.4	Test af homogenitetshypotese	221
5.5	Estimation af faste parametre	223
5.6	Estimation af de tilfældige effekter, BLUP-estimation	228
5.7	Lidt mere om hierariske modeller	229
5.7.1	Generelt	229
5.7.2	Aposteriorifordeling svarende til normal-normal fordeling	231
5.8	I SAS	234
5.8.1	Mixed	237
5.9	Empirisk Bayes estimation, BLUP estimation	241
5.10	Flere eksempler på hierarkisk variation for normalfordelingsmodeller	248
5.10.1	Analyse af blokforsøg	248
5.10.2	Varierende regressionslinier	250
5.10.3	Flerdimensional normalfordeling	253
5.11	Aposteriorifordelinger for flerdimensionale normalfordelinger	257
6	Generelle lineære mixede modeller	259
6.0	Introduktion	259
6.1	Formulering af den generelle lineære mixed model	259

6.1.1	Eksempel: Envejs model med tilfældig effekt	260
6.2	Estimation	261
6.2.1	Estimation af fixed effects og variansparametre	261
6.2.2	Estimation af de tilfældige effekter	262
6.2.3	Samtidig estimation af β og $\mathbf{u}$	263
6.2.4	Empirisk Bayes-fortolkning	264
6.3	SAS-proceduren PROC MIXED	265
6.3.1	Eksempel på brug af PROC MIXED	265
7	Hierarkisk variation for exponentielle dispersionsfamilier	268
7.0	Indledning, modellering af overdispersion	268
7.1	Hierarkisk Poisson Gamma model	269
7.1.1	Eksempel, Variation mellem episoder af tordenvejr ved Cape Kennedy	269
7.1.2	Formulering af hierarkisk model	270
7.1.3	SAS PROC GENMOD	273
7.1.4	Eksempel, tordenvejr igen	273
7.2	Hierarkisk binomial beta model	274
7.2.1	Data	274
7.2.2	Formulering af hierarkisk model	276
7.3	Hierarkisk model for empiriske varianser for normaltfordelte variable	277
7.3.1	Den systematiske model	277
7.3.2	Den tilfældige model	278
7.3.3	Fortolkning af parametre i strukturfordelingen af σ^2	279
7.3.4	Marginal fordeling af stikprøvevariansen	280
7.4	Normalfordelingsmodeller med tilfældigt varierende varians.	284
7.5	Generel formulering af hierarkiske modeller for exponentielle dispersionsfamilier	285
7.6	Aposteriorifordeling	288
7.6.1	Generelle resultater vedrørende aposteriorifordelinger	289
7.7	Aposteriorifordeling for Poisson Gamma model	294
7.8	Normalfordelingen som miksturfordeling	295
8	sasdata	299
9	Diverse	305
9.1	om information	305

Forord

Nærværende notesæt er udarbejdet til brug for specialkursus ved IMM, 02423/

Dette specialkursus har været afholdt Notesættet

Ideen med noterne er Vi giver ikke så mange beviser. Ambitionen er snarere at skitsere nogle overordnede linier - og desuden at give lidt vejledning til praktisk brug.

For god ordens skyld erindres om, at denne teori hovedsageligt sigter mod problemer, der kan formuleres som modellering, estimation og test. Teorien er således ikke rettet mod behandling af, eksempelvis prædiktionsproblemer.

Rettet mod brug af SAS, fortrinsvis SAS Insight

referencer til standardstatistikbog, Shao [13] , Bickel og Doksum [3] m.fl.

Kapitel 1

Momenter og flerdimensionale stokastiske variable

1.0 Indledning

Dette afsnit tjener i det væsentlige som reference med angivelse af notation og egenskaber for momenter. Afsnit 1.3 angiver definitioner og resultater for momenter af flerdimensionale stokastiske variable, og endelig resumerer afsnit 1.4 tæthedsfunktion og momenter for en række almindeligt anvendte fordelinger.

1.1 Notation

Vi vil benytte en fælles notation for summer af sandsynlighedsfunktioner (for diskrete stokastiske variable) og for integraler af tæthedsfunktioner (for kontinuerte stokastiske variable), nemlig

$$\int h(y) f_Y(y) \nu\{dy\} = \begin{cases} \sum_i h(y_i) f_Y(y_i) & \text{for } Y \text{ diskret} \\ \int h(y) f_Y(y) dy & \text{for } Y \text{ kontinuert} \end{cases}$$

idet en sum kan fortolkes som et generaliseret integral (med hensyn til et diskret *mål*).

Vi vil bruge symbolerne $E[Y]$ og $V[Y]$ for at angive *middelværdi* og *varians* i

fordelingen af den stokastiske variable Y , nemlig

$$\begin{aligned} \mathbb{E}[Y] &= \int y f_Y(y) \nu\{dy\} \\ \mathbb{V}[Y] &= \int (y - \mathbb{E}[Y])^2 f_Y(y) \nu\{dy\} \end{aligned}$$

hvor $f_Y(y)$ angiver tæthedsfunktionen for Y .

1.2 Momenter

Definition 1.2.1 Betingede og marginale momenter

Lad X og Y være stokastiske variable med den simultane tæthed $f(x, y)$ (med hensyn til målet $\mu \times \nu$ på $\mathcal{X} \times \mathcal{Y}$)

Ved den marginale middelværdi af $\phi(X)$ vil vi forstå værdien af integralet

$$\mathbb{E}[\phi(X)] = \int_{\mathcal{X} \times \mathcal{Y}} \phi(x) f(x, y) \mu\{dx\} \nu\{dy\} = \int_{\mathcal{X}} \phi(x) k_X(x) \mu\{dx\}$$

hvor $k_X(x)$ angiver den marginale tæthed for X .

Ved den betingede middelværdi af $\phi(X)$ for givet $Y = y$ vil vi forstå den stokastiske variabel, der for $Y = y$ antager værdien

$$\int_{\mathcal{X}} \phi(x) h(x|y) \mu\{dx\}$$

Kort skriver vi

$$\mathbb{E}[\phi(X)|Y = y] = \int_{\mathcal{X}} \phi(x) h(x|y) \mu\{dx\} \quad (1.1)$$

Sætning 1.2.1 Relation mellem marginale og betingede momenter

Lad X , Y og Z være endimensionale stokastiske variable. Da gælder:

$\begin{aligned} \mathbb{E}[X] &= \mathbb{E}_Y[\mathbb{E}[X Y]] \\ \mathbb{V}[X] &= \mathbb{E}_Y[\mathbb{V}[X Y]] + \mathbb{V}_Y[\mathbb{E}[X Y]] \\ \text{COV}[X, Z] &= \mathbb{E}_Y[\text{COV}[X, Z] Y] + \text{COV}_Y[\mathbb{E}[X Y], \mathbb{E}[Z Y]] \end{aligned} \quad (1.2)$

Bevis:

Sætningen vises ved opstilling af integralerne og brug af en variant af Fubini's sætning.

▽

Bemærkning 1 Fortolkning ved tyngdepunkter og inertimomenter

Opfatter man forventningsværdien af en stokastisk variabel som *tyngdepunktet* i en massefordeling svarende til fordelingen af sandsynlighedsmassen ser man, at sætningen vedrørende den marginale forventningsværdi er analog til den fra Mekniken kendte sætning om bestemmelse af tyngdepunktet for et sammensat legeme. Opfatter man tilsvarende variansen for en stokastisk variabel som *inertimomentet* omkring (en akse gennem) tyngdepunktet ser man, at sætningen vedrørende den marginale varians svarer til at inertimomentet for et sammensat legeme bestemmes som summen af inertimomenterne omkring de lokale tyngdepunkter plus inertimomentet om det fælles tyngdepunkt med de lokale masser samlet i deres tyngdepunkt (Steiner's sætning).

▽

Sætning 1.2.2 Approksimative udtryk for middelværdi og varians for funktion af stok.var.

Lad X være en stokastisk variabel med middelværdi μ_X og varians σ_X^2 og betragt den stokastiske variable Y bestemt ved $Y = h(X)$, hvor $h(\cdot)$ er en differentiabel funktion. Da gælder

$$\begin{aligned} \mathbb{E}[Y] &= \mathbb{E}[h(X)] \simeq h(\mu_X) \\ \mathbb{V}[Y] &= \mathbb{V}[h(X)] \simeq \left(h'(\mu_X)\right)^2 \sigma_X^2 \end{aligned} \tag{1.3}$$

Bevis:

Sætningen vises ved betragtning af Taylor udviklingen

$$Y \simeq h(\mu_X) + (X - \mu_X)h'(\mu_X)$$

▽

Tilsvarende fås for en funktion $h(\cdot)$ af flere stokastiske variable $X_1, X_2, \dots, X_n$

$$Y = h(X_1, X_2, \dots, X_n)$$

ved Taylor-udvikling

$$\begin{aligned} Y &= h(X_1, X_2, \dots, X_n) \\ &\simeq h(\mu_1, \dots, \mu_n) + (X_1 - \mu_1) \frac{\partial h}{\partial x_1} + \dots + (X_n - \mu_n) \frac{\partial h}{\partial x_n} \end{aligned}$$

at

$$\begin{aligned} \mathbb{E}[Y] &\simeq h(\mu_1, \mu_2, \dots, \mu_n) \\ \mathbb{V}[Y] &\simeq \left(\frac{\partial h}{\partial x_1} \right)^2 \mathbb{V}[X_1] + \dots + \left(\frac{\partial h}{\partial x_n} \right)^2 \mathbb{V}[X_n] \\ &\quad + 2 \frac{\partial h}{\partial x_1} \frac{\partial h}{\partial x_2} \text{COV}[X_1, X_2] \\ &\quad + \dots + 2 \frac{\partial h}{\partial x_{n-1}} \frac{\partial h}{\partial x_n} \text{COV}[X_{n-1}, X_n] \end{aligned}$$

Kaldes ofte *fejlophobningsloven* (eng. *Law of error propagation*).

1.3 Momenter for flerdimensionale variable

For en p -dimensional stokastisk variabel

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix}$$

definerer vi forventningsværdien $\mathbb{E}[X]$ som vektoren

$$\mathbb{E}[X] = \begin{pmatrix} \mathbb{E}[X_1] \\ \mathbb{E}[X_2] \\ \vdots \\ \mathbb{E}[X_p] \end{pmatrix} \quad (1.4)$$

Forventningsværdien er lineær, dvs. såfremt X og Y er p -dimensionale stokastiske vektorer, da er

$$\begin{aligned} E[X + Y] &= E[X] + E[Y] \\ E[aX] &= aE[X] \end{aligned}$$

Såfremt $\mathbf{A}$ er en $n \times p$ matrix af konstanter, da gælder

$$E[\mathbf{A}X] = \mathbf{A} E[X]$$

Definition 1.3.1 Kovarians mellem flerdimensionale observationer

Kovariansen $\text{COV}[X, Y]$ mellem den p -dimensionale stokastiske vektor X og den q -dimensionale stokastiske vektor Y defineres som

$$\text{COV}[X, Y] = E[(X - E[X])(Y - E[Y])'] \quad (1.5)$$

Udtrykt ved kovarianserne for de endimensionale variable $X_1, \dots, X_p$ og $Y_1, \dots, Y_q$ har vi

$$\text{COV}[X, Y] = \begin{pmatrix} \text{COV}[X_1, Y_1] & \dots & \text{COV}[X_1, Y_q] \\ \text{COV}[X_2, Y_1] & \dots & \text{COV}[X_2, Y_q] \\ \vdots & \ddots & \vdots \\ \text{COV}[X_p, Y_1] & \dots & \text{COV}[X_p, Y_q] \end{pmatrix}$$

Der gælder

Sætning 1.3.1 Kovariansen er en symmetrisk, bilinear form

Kovariansen er lineær i hvert af sine to argumenter: Lad X og Y være p -dimensionale og lad U og V være q -dimensionale stokastiske vektorer. Da gælder

$$\begin{aligned} \text{COV}[X + Y, U] &= \text{COV}[X, U] + \text{COV}[Y, U] \\ \text{COV}[X, U + V] &= \text{COV}[X, U] + \text{COV}[X, V] \\ \text{COV}[aX, U] &= a\text{COV}[X, U] \end{aligned}$$

Såfremt $\mathbf{A}$ er en $n \times p$ og $\mathbf{B}$ en $m \times q$ matrix af konstanter gælder

$$\text{COV}[\mathbf{A}X, \mathbf{B}U] = \mathbf{A}\text{COV}[X, U]\mathbf{B}' \quad (1.6)$$

Ydermere gælder symmetrirelationen

$$\text{COV}[X, U] = (\text{COV}[U, X])'$$

▽

Definition 1.3.2 Dispersionsmatricen for en flerdimensional stokastisk variabel

For en p -dimensional stokastisk variabel X indfører vi dispersionsmatricen $D[X]$ ved

$$D[X] = \text{COV}[X, X] = \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])'] \quad (1.7)$$

Udtrykt ved momenterne for de endimensionale variable $X_1, X_2, \dots, X_n$ har vi

$$D[X] = \begin{pmatrix} V[X_1] & \text{COV}[X_1, X_2] & \dots & \text{COV}[X_1, X_p] \\ \text{COV}[X_2, X_1] & V[X_2] & \dots & \text{COV}[X_2, X_p] \\ \vdots & \vdots & \ddots & \vdots \\ \text{COV}[X_p, X_1] & \text{COV}[X_p, X_2] & \dots & V[X_p] \end{pmatrix}$$

Såfremt $\mathbf{A}$ er en $n \times p$ matrix af konstanter følger det af egenskaberne for kovariansen at

$$D[\mathbf{A}X] = \mathbf{A} D[X] \mathbf{A}' \quad (1.8)$$

endvidere har vi, såfremt X og Y begge er p -dimensionale stokastiske variable:

$$D[X + Y] = D[X] + D[Y] + \text{COV}[X, Y] + \text{COV}[Y, X]$$

▽

Sætning 1.3.2 Dispersionsmatricen er ikke-negativ definit

Lad X angive en p -dimensional stokastisk variabel. Da gælder at dispersionsmatricen $D[X]$ er en ikke-negativ definit symmetrisk matrix.

▽

For flerdimensionale variable X , Y og Z gælder de analoge relationer til (1.2):

$$\begin{aligned} E[X] &= E_Y[E[X|Y]] \\ D[X] &= E_Y[D[X|Y]] + D_Y[E[X|Y]] \\ \text{COV}[X, Z] &= E_Y[\text{COV}[X, Z|Y]] + \text{COV}_Y[E[X|Y], E[Z|Y]] \end{aligned} \quad (1.9)$$

Definition 1.3.3 Partiel varians og kovarians

Lad X , Y og Z angive (evt. flerdimensionale) stokastiske variable sådan at $D[X]$ har fuld rang.

Ved den partielle kovarians af Y og Z givet X vil vi forstå den bilineære form (bilineær i Y og Z)

$$\text{COV}[Y, Z : X] = \text{COV}[Y, Z] - \text{COV}[Y, X]D[X]^{-1}\text{COV}[X, Z] \quad (1.10)$$

Ved den partielle varians af Y givet X vil vi specielt forstå

$$D[Y : X] = \text{COV}[Y, Y : X] = D[Y] - \text{COV}[Y, X]D[X]^{-1}\text{COV}[X, Y] \quad (1.11)$$

For endimensionale variable X og Y finder vi altså specielt

$$V[Y : X] = V[Y] - \text{COV}[Y, X]^2/V[X]$$

Den partielle varians af Y givet X er *Schur komplementet* til $D[X]$ i

$$D \begin{bmatrix} X \\ Y \end{bmatrix}$$

Vi bemærker, at mens den betingede kovarians $\text{COV}[Y, Z|X = x]$ afhænger af den aktuelle værdi x af X , da afhænger den partielle kovarians $\text{COV}[Y, Z : X]$ kun af momenterne i fordelingen af X m.v.

Sætning 1.3.3 Den partielle kovarians $\text{COV}[Y, Z : X]$ er invariant overfor lineære transformationer af X .

Lad X være en p -dimensional variabel, og lad $\mathbf{A}$ være en $p \times p$ matrix af konstanter sådan at $\mathbf{A}$ har fuld rang.

Da gælder

$$\text{COV}[Y, Z : \mathbf{A}X] = \text{COV}[Y, Z : X] \quad (1.12)$$

▽

1.4 Fordelinger afledt af normalfordelingen

1.4.1 The log-normal distribution model

Definition and basic properties

The log-normal distribution model is used quite extensively to model the distribution of highly skewed, positive data, like distributions of income, growth, etc. For an example, see e.g. p. 207 in the textbook.

The log-normal distribution model has the property that when the distribution model for the raw data is a log-normal distribution, then the logarithmically transformed data follows a normal distribution.

The probability density function for a log-normal distribution model is given by

$$p_Y(y) = \frac{1}{y\beta\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\ln(y) - \alpha}{\beta}\right)^2\right] \quad \text{for } y > 0$$

and $p_Y(y) = 0$ otherwise, and with $\beta > 0$, and α denoting *parameters* of the distribution model.

For short we write $Y \sim \text{LN}(\alpha, \beta^2)$ to denote that the random variable Y is distributed according to the logarithmic normal distribution model with density given above.

Mean and variance

Let $Y \sim \text{LN}(\alpha, \beta^2)$, then the mean and variance of Y are given by

$$\begin{aligned}\mu_Y &= \text{E}[Y] = \exp\left(\alpha + \frac{1}{2}\beta^2\right) \\ \sigma_Y^2 &= \text{V}[Y] = \mu_Y^2 \left(\exp(\beta^2) - 1\right)\end{aligned}$$

Thus, the *coefficient of variation* (relative standard deviation) for the distribution of Y is

$$\gamma_Y = \frac{\sigma_Y}{\mu_Y} = \sqrt{\exp(\beta^2) - 1}$$

a dimensionless quantity that does not depend on the parameter α .

Relation to the normal distribution model

When $Y \sim \text{LN}(\alpha, \beta^2)$, and we put $Z = \ln(Y)$ then $Z \sim \text{N}(\alpha, \beta^2)$, i.e., $Z = \ln(Y)$ follows a normal distribution model with mean $\mu_Z = \alpha$, and variance, $\sigma_Z^2 = \beta^2$. (Thus, α and β are expressed in logarithms of the original data units).

Note on transformations

You will notice from the expressions above that although $Y = \exp(Z)$, then we have

$$\mathbb{E}[Y] = \exp(\mu_Z) \cdot \exp\left(\frac{1}{2}\sigma_Z^2\right)$$

which is *greater than* $\exp(\mu_Z)$. Thus, the result on the mean value of a linear transformation of a random variable p. 160 in the textbook is not generally valid for *nonlinear transformations*.

Cumulative distribution function

When $Y \sim \text{LN}(\alpha, \beta^2)$, the cumulative distribution function is

$$F_Y(y) = \Phi\left(\frac{\ln(y) - \alpha}{\beta}\right)$$

with $\Phi(\cdot)$ denoting the cumulative distribution function for the standardized normal distribution. Thus, all probabilities related to log-normal distribution models may be determined from the standardized normal distribution.

1.4.2 The χ^2 -distribution model

Definition and basic properties

The family of χ^2 -distribution models is introduced briefly in the textbook in connection with tests related to categorical variables on page 408. However, the use of the χ^2 -distribution model is much wider than indicated in that section of the textbook, and we shall therefore for later reference provide a summary of the properties of χ^2 -distribution models.

The family of χ^2 -distribution models is a family of distribution models on the positive real line with probability density function

$$p_Y(y) = \frac{1}{2\Gamma(f/2)} \left(\frac{y}{2}\right)^{f/2-1} \exp(-y/2) \quad \text{for } 0 < y$$

and $p_Y(y) = 0$ otherwise, and with $\Gamma(\cdot)$ denoting the *Gamma*-function, (see 1.4.6 below).

The symbol f denotes a positive integer-valued *parameter* that determines which of the many χ^2 -distributions that is considered. The parameter f is termed the *degrees of freedom*) for the distribution. In applications the degrees of freedom are usually known.

A random variable that is distributed according to a χ^2 -distribution model with f degrees of freedom is often denoted by the symbol χ_f^2 .

The χ^2 -distribution model is a special case of a larger class of distribution models, the so-called *gamma*-distribution models. For $Y \sim \chi_f^2$ one has $Y \sim G(\alpha, \beta)$ with $\alpha = f/2$ and $\beta = 2$.

Mean and variance

Let $Y \sim \chi_f^2$, then the mean and variance of Y are given by

$$E[Y] = f; \quad V[Y] = 2f \quad (1.13)$$

1.4.3 Relation to the normal distribution model

The χ^2 -distribution model is often encountered as distribution model for the variance, S^2 , in the data distribution in normal distribution models, as indicated by the following theorems.

Sætning 1.4.1 *Kvadratet på en standardiseret normalfordelt størrelse følger en χ_1^2 -fordeling* Let $Y \sim N(0, 1)$, and let $Z = Y^2$. Then

$$Z \sim \chi_1^2$$

Bevis:

The proof of the theorem relies upon results on distribution models for transformations of random variables. ∇

Observe that as a consequence of this theorem, the following assertion holds: when $Y \sim N(\mu, \sigma^2)$, then

$$\frac{(Y - \mu)^2}{\sigma^2} \sim \chi_1^2 \quad (1.14)$$

Sætning 1.4.2 *Fordeling af kvadratsummer af standardiseret normalfordelte størrelser*

Let $Y_1, Y_2, \dots, Y_n$ be independent random variables, each distributed according to the standard normal distribution model, $Y_i \sim N(0, 1)$, and let

$$\begin{aligned} \text{USS} &= \sum_{i=1}^n Y_i^2 \\ \text{CSS} &= \sum_{i=1}^n (Y_i - \bar{Y})^2 \end{aligned}$$

denote the raw (uncorrected) sum of squares, and the sum of squared deviations from the mean (observations corrected by the mean), respectively. Then

1. $\text{USS} \sim \chi_n^2$
2. $\text{CSS} \sim \chi_{n-1}^2$
3. $n\bar{Y}^2 \sim \chi_1^2$
4. $\text{USS} = \text{CSS} + n\bar{Y}^2$, and CSS and $n\bar{Y}^2$ are independent random variables

∇

- o -

Note that the decomposition

$$\text{USS} = \text{CSS} + n\bar{Y}^2$$

is nothing but the decomposition

$$\sum_{i=1}^n Y_i^2 = \sum_{i=1}^n (Y_i - \bar{Y})^2 + n\bar{Y}^2$$

that was considered in the supplementary note on partitioning sums of squares and degrees of freedom. (It may be shown that orthogonality of subspaces is equivalent to independence of the corresponding sums of squares)

It is not surprising that $n\bar{Y}^2 \sim \chi_1^2$, because it follows from the properties of the normal distribution that $\bar{Y} \sim N(0, 1/n)$, and then this result follows from (1.14).

The result above may be used to demonstrate that when $Y_1, Y_2, \dots, Y_n$ are independent normally distributed random variables with the same mean and variance, i.e., $Y_i \sim N(\mu, \sigma^2)$, and S^2 denotes the variance of the data-distribution for these n observations,

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2 \quad (1.15)$$

then the mean of the distribution model for S^2 is

$$\mu_{S^2} = \sigma^2$$

(cfn. (1.13)). This result is sometimes used as an argument for dividing the sum of squared deviations, CSS by its *degrees of freedom*, $(n-1)$, rather by the number of observations, n .

Further one finds from (1.13) that the variance of the distribution model for S^2 is

$$\sigma_{S^2}^2 = \frac{2(\sigma^2)^2}{n-1}$$

1.4.4 The t-distribution model

Definition and basic properties

The t-distribution model is introduced briefly in the textbook on p. 248. Below we shall summarize some properties of this distribution model for later reference.

The family of t-distribution models is a family of distribution models on the real line with probability density function

$$p_t(t) = \frac{\Gamma((f+1)/2)}{\sqrt{f\pi}\Gamma(f/2)} \left(1 + \frac{t^2}{f}\right)^{-(f+1)/2}$$

with $\Gamma(\cdot)$ denoting the *Gamma*-function, (see 1.4.6 below).

The symbol f denotes a positive integer-valued *parameter* that determines which of the many t-distributions that is considered. The parameter f is termed the *degrees of freedom*) for the distribution. In applications the degrees of freedom are usually known.

A random variable that is distributed according to a t-distribution model with f degrees of freedom is often denoted by the symbol t_f

Mean and variance

Let $Z \sim t_f$.

When $f > 1$ the mean of the distribution exists, and

$$E[Z] = 0 \quad \text{for } f > 1$$

When $f > 2$ the variance exists, and

$$V[Z] = \frac{f}{f-2} \quad \text{for } f > 2$$

Sætning 1.4.3 *Relation to the normal distribution model*

Let $U \sim N(0, 1)$, and let $Z \sim \chi_f^2$ be *independent* random variables, and consider the random variable

$$T = \frac{U}{\sqrt{Z/f}}$$

Then $T \sim t_f$, i.e. a t-distribution model with f degrees of freedom.

▽

This result is utilized when making statistical inference in normal distribution models:

Let $Y_1, Y_2, \dots, Y_n$ be independent normally distributed random variables with the same mean and variance, i.e., $Y_i \sim N(\mu, \sigma^2)$, and let as usual $\bar{Y}$ denote

the mean, and S^2 denote the variance (1.15) in the data distribution. It has already been observed that

$$\begin{aligned} U &= \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \\ Z &= f \cdot S^2/\sigma^2 \sim \chi_f^2 \end{aligned}$$

with $f = n - 1$, and with U and Z being independent (as asserted in section 1.4.3 above).

Now consider

$$T = \frac{\bar{Y} - \mu}{S/\sqrt{n}} \tag{1.16}$$

and note that T may be expressed as

$$T = \frac{\bar{Y} - \mu}{S/\sqrt{n}} = \frac{(\bar{Y} - \mu)/(\sigma/\sqrt{n})}{\sqrt{S^2/\sigma^2}} = \frac{U}{\sqrt{Z/f}}$$

It then follows from the theorem above that $T \sim t_f$.

Small and large sample properties

When comparing U and T , you note that the difference between these two variables is that the denominator in U is the *known* value of the standard deviation $\sigma/\sqrt{n}$ for $\bar{Y}$, whereas the denominator in T has replaced this value by an *estimate* of the standard deviation, $\hat{\sigma}(\bar{Y}) = S/\sqrt{n}$. Consequently, the distribution of T will have heavier tails than the normal distribution of U , but when the degrees of freedom, f , increases, the estimate, $\hat{\sigma}(\bar{Y})$ becomes more precise, and hence the distribution of T will approach a standard normal distribution. Therefore, the normal distribution may be used to approximate the t-distribution when the degrees of freedom are large, e.g. greater than 30

Degrees of freedom and sample size for mean of data

Also note from the relation

$$T = \frac{\bar{Y} - \mu}{\hat{\sigma}(\bar{Y})} = \frac{\bar{Y} - \mu}{S/\sqrt{n}} \sim t_f \tag{1.17}$$

that the *degrees of freedom*, f , refers to the *estimate for σ^2* , whereas n refers to the number of observations that are used for determining the mean $\bar{Y}$.

This distinction is useful when the variance estimate, S^2 , has been obtained from other sources than the estimate, $\bar{Y}$, e.g. when *pooling* estimators of variance (cfn. the textbook p. 262).

1.4.5 The F-distribution model

Definition and basic properties

The F-distribution model is introduced briefly in the textbook on p. 477 and 529. Below we shall summarize some properties of this distribution model for later reference.

The family of F-distribution models is a family of distribution models on the positive real line with probability density function

$$p_X(x) = \frac{\Gamma((n+m)/2)}{\Gamma(n/2)\Gamma(m/2)} \left(\frac{n}{m}\right)^{n/2} \frac{x^{n/2-1}}{(1+nx/m)^{(n+m)/2}} \quad \text{for } 0 < x < \infty$$

with $\Gamma(\cdot)$ denoting the *Gamma*-function.

The symbols n og m denotes positive integer-valued *parameters* that determine which of the many F-distributions that is considered. The parameters are termed the *degrees of freedom*) for the distribution. In applications the degrees of freedom are usually known.

A random variable that is distributed according to a F-distribution model with (n, m) degrees of freedom is often denoted by the symbol $F_{n,m}$.

Sometimes the following relation between percentiles of the F-distribution may be useful

$$F_{n,m;q} = \frac{1}{F_{m,n;1-q}} \quad (1.18)$$

Mean and variance

Let $Y \sim F_{n,m}$.

When $m > 2$ the mean of the distribution exists, and

$$E[Y] = \frac{m}{m-2} \quad \text{for } m > 2$$

When $m > 4$ the variance exists, and

$$V[Y] = \frac{m^2(2n+2m-4)}{(m-2)^2(m-4)n} \quad \text{for } m > 4$$

Sætning 1.4.4 *Relation to the t-distribution model* Let $Y \sim t_f$, and let $Z = Y^2$. Then

$$Z \sim F_{1,f}$$

i.e. the square of a t_f -distributed variable follows a F-distribution model with $(1, f)$ degrees of freedom. ▽

When considering linear normal distribution models, the following result is often used:

Sætning 1.4.5 *Relation to the normal distribution model* Let $Z_1 \sim \chi_n^2$, and $Z_2 \sim \chi_m^2$ be independent, and consider

$$Y = \frac{Z_1/n}{Z_2/m}$$

then $Y \sim F_{n,m}$.

Thus, the F-distribution model describes the distribution of the ratio between two *independent* estimators of variance under the assumption that these estimators both reflect the *same* variance, σ^2 .

Accordingly, the parameter, n , is often denoted the degrees of freedom for the *numerator*, and m the degrees of freedom for the *denominator*. ▽

1.4.6 The Gamma-function

The gamma-function, $\Gamma(\cdot)$, is a function of a real positive argument, given by

$$\Gamma(x) = \int_0^\infty t^{x-1} \exp(-t) dx$$

The function satisfies the recursive relation

$$x\Gamma(x) = \Gamma(x+1)$$

In particular, $\Gamma(1) = 1$ and $\Gamma(1/2) = \sqrt{\pi}$.

For integer values, n , one therefore has

$$\begin{aligned}\Gamma(n) &= 1 \cdot 2 \cdot 3 \cdots (n-1) = (n-1)! \\ \Gamma(n+1/2) &= \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{2^n} \sqrt{\pi}\end{aligned}$$

1.4.7 Oversigt over en række sædvanlige fordelinger

Tabel 1.1: Frekvensfunktion, middelværdi og varians for en række sædvanlige fordelinger

Navn	Støtte	Frekvensfunkt	$E[Y]$	$V[Y]$
Binomial $\text{Bin}(n, p)$	$0, 1, 2, \dots, n$ $n \in \mathbf{N}, p \in [0, 1]$	$\binom{n}{y} p^y (1-p)^{n-y}$	np	$np(1-p)$
Poisson $P(\lambda)$	$0, 1, 2, \dots$ $\lambda \in \mathbf{R}_+$	$\frac{\lambda^y}{y!} \exp(-\lambda)$	λ	λ
Neg Bin $\text{NB}(r, p)$	$0, 1, 2, \dots$ $r \in \mathbf{R}_+, p \in]0, 1]$	$\binom{r+y-1}{y} p^r (1-p)^y$	$\frac{r(1-p)}{p}$	$\frac{r(1-p)}{p^2}$
Normal $N(\mu, \sigma^2)$	$\mathbf{R}$ $\mu \in \mathbf{R}, \sigma^2 \in \mathbf{R}_+$	$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$	μ	σ^2
Gamma $G(\alpha, \beta)$	$\mathbf{R}_+$ $\alpha \in \mathbf{R}_+, \beta \in \mathbf{R}_+$	$\frac{1}{\Gamma(\alpha)\beta} \left(\frac{y}{\beta}\right)^{\alpha-1} \exp(-y/\beta)$	$\alpha\beta$	$\alpha\beta^2$

Kapitel 2

Likelihoodfunktion

2.0 Introduktion og oversigt

I dette afsnit introduceres begrebet *likelihoodfunktion* svarende til et forelagt observationssæt under en statistisk model. Likelihoodfunktionen er en funktion af modellens *parametre*, der udtrykker graden af overensstemmelse mellem data og en parameter værdi.

Endvidere indføres *scorefunktionen* som er den afledede af logaritmen til likelihoodfunktionen, og den *observerede information*, som er den anden afledede, dvs krumningen af log-likelihood'en, og sluttelig anføres nogle vigtige resultater vedrørende scorefunktion og information for transformationer af parametre.

Endelig indeholder afsnit 2.2 nogle simple eksempler til illustration af de lidt abstrakte begreber.

The likelihood-function and the maximum-likelihood principle are discussed in most textbooks on mathematical statistics. A recent, rather general reference, exploring many contemporary examples of likelihood inference is the textbook by Y.Pawitan [24].

2.1 Likelihoodfunktion, Scorefunktion og Informationsmatrix

Sandsynlighedsmodel og statistisk model

En *sandsynlighedsmodel* for et observationssæt består af

- et *udfaldsrum* (sample space),
- et *hændelsessystem* (events), $\mathcal{A}$
- et *sandsynlighedsmål*, (probability measure) P , på $\mathcal{A}$

En *statistisk model* er en *parametriseret*¹ mængde af sandsynlighedsmodeller. Vi vil indskrænke os til at betragte modeller for stokastiske variable $Y_1, Y_2, \dots, Y_n$.

Definition 2.1.1 *Likelihoodfunktion, loglikelihoodfunktion*

Lad den statistiske model for observationssættet $Y_1, Y_2, \dots, Y_n$ være givet ved familien af simultane tætheder

$$\{f_Y(y_1, y_2, \dots, y_n; \theta)\}_{\theta \in \Theta} \quad (2.1)$$

med hensyn til et mål ν_n på $\mathcal{Y}^n \subset \mathbb{R}^n$

Ved likelihoodfunktionen for θ svarende til observationen $Y_1 = y_1, Y_2 = y_2, \dots, Y_n = y_n$ forstår vi en funktion

$$L(\theta; \mathbf{y}) = c(y_1, y_2, \dots, y_n) f_Y(y_1, y_2, \dots, y_n; \theta) \quad (2.2)$$

for $\theta \in \Theta$ og hvor $c(y_1, y_2, \dots, y_n)$ er en vilkår proportionalitetskonstant.

Likelihoodfunktionen angiver således den simultane tæthed for observationssættet som funktion af parameteren θ . Vi bemærker, at mens tætheden for observationssættet opfattes som en funktion af de mulige observationsværdier, $(y_1, y_2, \dots, y_n)$ for en fastholdt værdi af parameteren θ , så opfattes likelihoodfunktionen som en funktion af parameteren θ for en fastholdt værdi (den observerede) af observationssættet $(y_1, y_2, \dots, y_n)$. Vi har symboliseret

¹Parametriseret betyder blot at samlingen af sandsynlighedsmodeller er indekseret.

denne forskel ved at lade det første argument i likelihoodfunktionen være θ , mens det første argument i den simultane tæthed er et muligt observations-sæt.

Ofte er det mere bekvemt at betragte logaritmen til likelihoodfunktionen.

<p>Vi indfører derfor log-likelihoodfunktionen svarende til observationssættet $\mathbf{y}$ som</p> $l(\theta; \mathbf{y}) = \ln(L(\theta; \mathbf{y})) \quad (2.3)$

hvor $L(\theta; \mathbf{y})$ er givet ved (2.2).

Vi vil almindeligvis udelade afhængigheden af observationen $\mathbf{y}$ ved angivelsen af likelihoodfunktionen. ∇

Undertiden betragter man den såkaldte *profillikelihood* til behandling af situationer, hvor parameteren θ er flerdimensional, og hvor enkelte komponenter af θ betragtedes som “nuisance” parametre, dvs irrelevante for inferensproblemet, og som derfor maksimeres bort.

Definition 2.1.2 Scorefunktionen

Betragt en familie af tætheder (2.1), hvor parameterområdet Θ er en åben delmængde af $\mathbb{R}^k$. Såfremt log-likelihoodfunktionen er kontinuert differentiable, betragter vi vektorfunktionen

$l'_\theta(\theta; \mathbf{y}) = \frac{\partial}{\partial \theta} l(\theta; \mathbf{y}) = \begin{pmatrix} \frac{\partial}{\partial \theta_1} l(\theta; \mathbf{y}) \\ \vdots \\ \frac{\partial}{\partial \theta_k} l(\theta; \mathbf{y}) \end{pmatrix} \quad (2.4)$

Funktionen $l'_\theta(\theta; \mathbf{y})$ benævnes scorefunktionen. ∇

Da scorefunktionen afhænger af observationssættet $\mathbf{y}$, er den en stokastisk

størrelse. Såfremt det er tilladt at ombytte integration og differentiation, følger det ved differentiation af udtrykket

$$\int f_Y(\mathbf{y}; \theta) \mu\{dy\} = 1 \quad (2.5)$$

med hensyn til θ , at

$\mathbb{E}_\theta \left[\frac{\partial}{\partial \theta} l(\theta; \mathbf{Y}) \right] = \mathbf{0} \quad (2.6)$
--

Betingelsen herfor er essentielt, at støtten for fordelingerne ikke afhænger af θ .

Definition 2.1.3 Observeret og forventet information

Matricen

$$\mathbf{j}(\theta; \mathbf{y}) = - \frac{\partial^2}{\partial \theta \partial \theta'} l(\theta; \mathbf{y}) \quad (2.7)$$

med elementer

$$\mathbf{j}(\theta; \mathbf{y})_{ij} = - \frac{\partial^2}{\partial \theta_i \partial \theta_j} l(\theta; \mathbf{y})$$

kaldes for den observerede information svarende til observationen $\mathbf{y}$ og parameterværdien θ .

Forventningsværdien af den observerede information,

$$\mathbf{i}(\theta) = \mathbb{E}[\mathbf{j}(\theta; \mathbf{Y})] \quad (2.8)$$

med $\mathbf{j}_\theta(\theta; \mathbf{Y})$ givet ved (2.7) kaldes for den forventede information, eller blot informationsmatricen svarende til parameterværdien θ . Den forventede information (2.8) kaldes undertiden Fisher informationen. ∇

Bemærkning 1 Den forventede information er dispersionsmatricen for scorefunktionen

Ved differentiation af (2.5) to gange, finder man

$$\mathbb{E}_\theta \left[\frac{\partial^2}{\partial \theta \partial \theta'} l(\theta; \mathbf{Y}) \right] + \mathbb{E}_\theta \left[\frac{\partial}{\partial \theta} l(\theta; \mathbf{Y}) \left(\frac{\partial}{\partial \theta} l(\theta; \mathbf{Y}) \right)' \right] = \mathbf{0} ,$$

hvorfor

$$\begin{aligned} \mathbf{i}(\theta) &= \mathbb{E}_{\theta} \left[- \frac{\partial^2}{\partial \theta \partial \theta'} l(\theta; \mathbf{Y}) \right] = \mathbb{E}_{\theta} \left[\frac{\partial}{\partial \theta} l(\theta; \mathbf{Y}) \left(\frac{\partial}{\partial \theta} l(\theta; \mathbf{Y}) \right)' \right] \\ &= \mathbf{D}_{\theta}[l'_{\theta}(\theta; \mathbf{Y})] \end{aligned} \quad (2.9)$$

hvor symbolet $\mathbf{D}[\cdot]$ betyder "dispersion", dvs varians- kovarians matrix.

Da informationsmatricen således er en dispersionsmatrix, er den positiv semi-definit.

Den observerede information angiver den aktuelle krumning (med modsat fortegn) af log-likelihoodfunktionen svarende til observationen $\mathbf{y}$ og parameterværdien θ .

Informationsmatricen angiver tilsvarende den forventede krumning (med modsat fortegn) af log-likelihoodfunktionen svarende til parameterværdien θ . Informationsmatricen er således et mål for, hvor godt parameteren bestemmes. Hvis informationsmatricen er "stor", er parameteren godt bestemt. ∇

Definition 2.1.4 Likelihoodfunktion for transformationer af parameterrommet

Lad $\psi = \psi(\theta)$ være en enentydig afbildning af $\Omega \subset \mathbb{R}^k$ ind i $\Psi \subset \mathbb{R}^k$, der er vilkårligt ofte differentiabel begge veje. Parametriseringen ved ψ er således blot en alternativ parametrisering af modellen.

Vi har likelihood- og loglikelihoodfunktionen svarende til parametriseringen ved ψ

$$\begin{aligned} L_{\Psi}(\psi; \mathbf{y}) &= L_{\Omega}(\theta(\psi); \mathbf{y}) \\ \ell_{\Psi}(\psi; \mathbf{y}) &= \ell_{\Omega}(\theta(\psi); \mathbf{y}) \end{aligned} \quad (2.10)$$

∇

Likelihoodfunktionen afhænger ikke af den parametrisering, der er valgt for parameterrommet Ω .

Relationen (2.10) indebærer, at man ikke kan fortolke en likelihood (evt.

normeret) som en sandsynlighedsfordeling på Ω , (for så skulle Jacobianten jo også indgå i (2.10)). Likelihood'en kan opfattes som en funktion, der er defineret på den betragtede samling af sandsynlighedsmodeller, og som ikke afhænger af den valgte parametrisering.

Imidlertid afhænger de afledede af log-likelihood funktionen af parametriseringen.

Bemærkning 1 *Scorefunktion og informationsmatrix for transformerede parametre*

Betragt en anden parametrisering givet ved $\theta = \theta(\beta) \in \mathbb{R}^k$ for $\beta \in B \subset \mathbb{R}^m$ med $m \leq k$ og antag, at afbildningen er kontinuert differentiabel. Da er scorefunktionen svarende til parametersættet β bestemt ved

$$l'_\beta(\beta; \mathbf{y}) = \mathbf{J}' l'_\theta(\theta(\beta); \mathbf{y}) \quad (2.11)$$

hvor Jacobi-matricen $\mathbf{J}$ er givet ved

$$\mathbf{J} = \frac{\partial \theta}{\partial \beta}$$

med elementer

$$\mathbf{J}_{ij} = \frac{\partial \theta_i}{\partial \beta_j} \quad i = 1, 2, \dots, k; \quad j = 1, 2, \dots, m$$

Den observerede information fremkommer ved brug af reglerne for sammensatte funktioner. Den bliver lidt kompliceret, men den forventede information med hensyn til β er givet ved

$$\mathbf{i}_\beta(\beta) = -\mathbf{E} \left[\frac{\partial^2}{\partial \beta \partial \beta'} l(\theta(\beta); \mathbf{Y}) \right] = \mathbf{J}' \mathbf{i}_\theta(\theta(\beta)) \mathbf{J} \quad (2.12)$$

hvor Jacobi-matricen $\mathbf{J}$ er som ovenfor.

Resultatet vises ved at bemærke, at den forventede information kan udtrykkes som dispersionsmatricen for scorefunktionen (2.11). ∇

2.2 Eksempler

Eksempel 2.2.1 Likelihoodfunktion

Der er udvalgt tre kuglelejer tilfældigt fra en produktion af kuglelejer, og for hvert leje har man registreret afvigelsen af lejets diameter fra 185.0. Man fandt følgende værdier 4.6; 6.3 and 5.0

Man formulerer en **model**

- **Model:** C+E model, $Y = \mu + \epsilon$
- **Data:** $Y_i = \mu + \epsilon_i$
- **Antagelser:**
 - Y_1, Y_2, Y_3 er uafhængige
 - $Y_i \sim N(\mu, \sigma^2)$
 - σ^2 er kendt, $\sigma^2 = 1$,

Der er således kun én ukendt modelparameter, $\mu_Y = \mu$.

Den simultane tæthedsfunktion Y_1, Y_2, Y_3 er

$$f_{Y_1, Y_2, Y_3}(y_1, y_2, y_3; \mu) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(y_1 - \mu)^2}{2}\right] \times \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(y_2 - \mu)^2}{2}\right] \times \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(y_3 - \mu)^2}{2}\right] \quad (2.13)$$

som for enhver værdi af μ er en funktion af de tre variable y_1, y_2, y_3 . (Man kan forsøge at forestille sig en klokkeformet flade defineret på det tredimensionale rum).

Når vi har observeret **data værdierne**, $y_1 = 4.6$; $y_2 = 6.3$ and $y_3 = 5.0$, er likelihood-funktionen

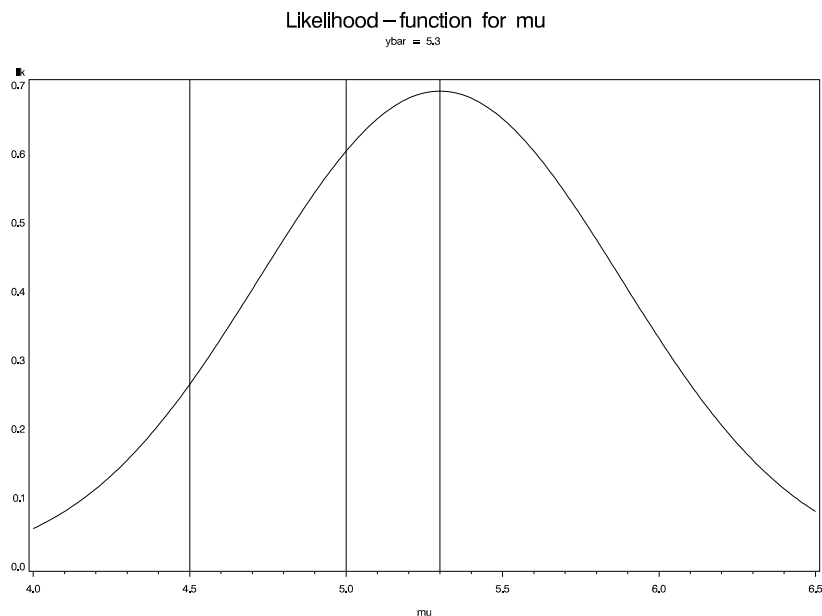
$$L_{4.6, 6.3, 5.0}(\mu) = f_{Y_1, Y_2, Y_3}(4.6, 6.3, 5.0; \mu) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(4.6 - \mu)^2}{2}\right] \times \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(6.3 - \mu)^2}{2}\right] \times \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(5.0 - \mu)^2}{2}\right]$$

kun funktion af μ . Bemærk, at likelihood-funktionen udtrykker den infinitesimale sandsynlighed for at få stikprøveresultatet (4.6, 6.3, 5.0) som funktion af den ukendte parameter μ .

Ved reduktion af likelihood-funktionen får man

$$L_{4.6, 6.3, 5.0}(\mu) = \frac{1}{(\sqrt{2\pi})^3} \exp\left[-\frac{1.58^2}{2}\right] \exp\left[-\frac{3(5.3 - \mu)^2}{2}\right] \quad (2.14)$$

Likelihoodfunktionen er skitseret i nedenstående graf



▽

Bemærk, at sandsynlighedstætheden (2.13) er en funktion af (y_1, y_2, y_3) , der beskriver *variationen* i fordelingen af potentielle data, mens likelihood-funktionen (2.14) er en funktion af μ , der beskriver den relative *plausabilitet* af de forskellige værdier af μ i lys af de givne data. Likelihoodfunktionen angiver i hvor høj grad forskellige værdier af μ stemmer overens med de observerede data. En værdi af μ , der giver anledning til en stor værdi af likelihoodfunktionen, er i bedre overensstemmelse med data, end en værdi af μ , der giver anledning til en mindre værdi af likelihoodfunktionen.

Sufficiency of the sample mean

In the example above, we used the numeric data values $y_1 = 4.6$; $y_2 = 6.3$ and $y_3 = 5.0$. However, one obtains more general insight if we just use the symbols

(y_1, y_2, y_3) for the data values. Using this notation, the likelihood-function is

$$\begin{aligned}
L_{y_1, y_2, y_3}(\mu) &= f_{Y_1, Y_2, Y_3}(y_1, y_2, y_3; \mu) = \\
&= \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(y_1 - \mu)^2}{2}\right] \times \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(y_2 - \mu)^2}{2}\right] \times \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(y_3 - \mu)^2}{2}\right] \\
&= \frac{1}{(\sqrt{2\pi})^3} \exp\left[-\frac{\sum (y_i - \bar{y})^2}{2}\right] \exp\left[-\frac{3(\bar{y} - \mu)^2}{2}\right]
\end{aligned} \tag{2.15}$$

which shows that (except for a factor not depending on μ), the likelihood-function does only depend on the observations (y_1, y_2, y_3) through the average $\bar{y} = \sum y_i/3$. Thus, whatever be the individual values (y_1, y_2, y_3) underlying a given value of $\bar{y}$, they give rise to the same likelihood-function.

Therefore, as long as we are concerned with inference about μ under this normal distribution model, it is sufficient to consider the sample mean $\bar{y} = \sum y_i/3$, and disregard the individual values (y_1, y_2, y_3) . We say that the sample mean, $\bar{Y} = \sum Y_i/3$ is a *sufficient statistic* for the parameter μ .

In general, a (possibly vector-valued) function $t(Y_1, Y_2, \dots, Y_n)$ is said to be a sufficient statistic for a (possibly vector-valued) parameter, θ , if the probability density function (or probability mass function) for $t(Y_1, Y_2, \dots, Y_n)$ may be factorized into a product

$$p_{Y_1, \dots, Y_n}(y_1, \dots, y_n; \theta) = h(y_1, \dots, y_n)g(t(y_1, y_2, \dots, y_n); \theta)$$

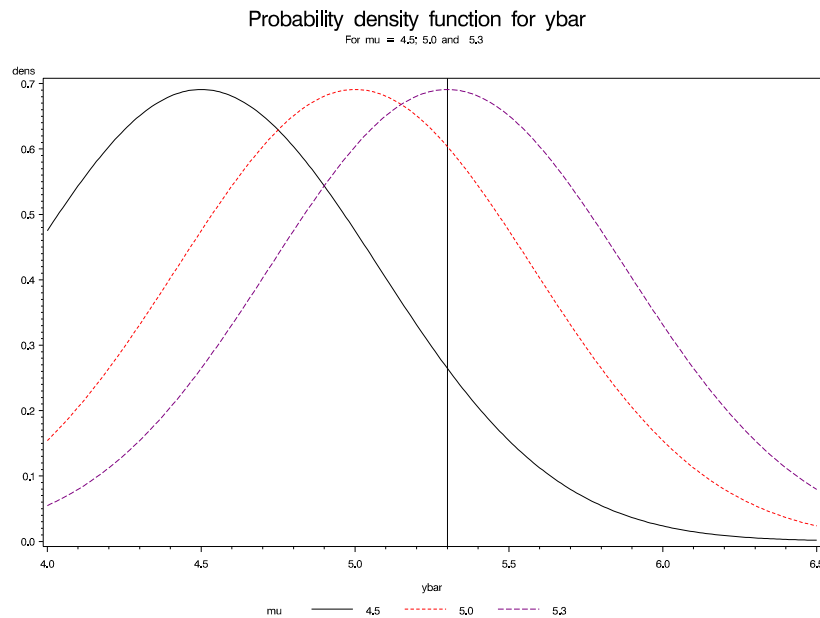
with the factor $h(y_1, \dots, y_n)$ not depending on the parameter θ , and the factor $g(t(y_1, y_2, \dots, y_n); \theta)$ only depending on $y_1, \dots, y_n$ through the function $(t(\cdot, \dots, \cdot))$. Thus, if we know the value of $t(y_1, y_2, \dots, y_n)$, the individual values $y_1, \dots, y_n$ do not contain further information about the value of μ .

In the example above, it was found that under the normal distribution model, the sample mean contains all relevant information about the population mean μ . (The ideas of robustness and outlierresistance refer to situations where one is not convinced that the distribution model holds).

Example (continued)

To get a better feeling for the likelihood-function, consider again the example above. Under the specified normal distribution model, the distribution of the sample average $\bar{Y}$ is a $N(\mu, 1/3)$ -distribution. The graph below shows the probability density function for $\bar{Y}$ when $\mu = 4.5$, $\mu = 5.0$, and $\mu = 5.3$, respectively.

When considering the likelihood-function corresponding to $\bar{y} = 5.3$, only the values at $\bar{y} = 5.3$ are used (indicated by a vertical line at $\bar{y} = 5.3$). In the previous graph of the likelihood-function these values are indicated by vertical bars at $\mu = 4.5$, $\mu = 5.0$, and $\mu = 5.3$. Thus, the likelihood-function is obtained by sliding the graph of the probability density function for μ through all possible values of μ , and plotting the value corresponding to $\bar{y} = 5.3$.



Example (scorefunktion)

Af (2.14) finder man logaritmen til likelihoodfunktionen

$$l(\mu; 4.6, 6.3, 5.0) = - \frac{3(5.3 - \mu)^2}{2} + C(4.6, 6.3, 5.0)$$

hvorfor scorefunktionen er

$$l'_\mu(\mu; 4.6, 6.3, 5.0) = 3 \cdot (5.3 - \mu)$$

og den observerede information er

$$\mathbf{j}(\mu; 4.6, 6.3, 5.0) = 3$$

For at bestemme den forventede information bliver vi nødt til at foretage de analoge regninger med brug af symboler Y_1, Y_2, Y_3 for data.

Fra (2.15) finder man

$$l(\mu; Y_1, Y_2, Y_3) = -\frac{3(\bar{Y} - \mu)^2}{2} - 3 \ln(\sqrt{2\pi}) - \frac{\sum (Y_i - \bar{Y})^2}{2}$$

hvorfor scorefunktionen bliver

$$l'_\mu(\mu; Y_1, Y_2, Y_3) = 3(\bar{Y} - \mu)$$

og den observerede information er

$$\mathbf{j}(\mu; Y_1, Y_2, Y_3) = 3$$

der ikke afhænger af observationerne Y_1, Y_2, Y_3 , hvorfor den forventede information ligeledes er

$$\mathbf{i}(\mu) = \mathbb{E}[\mathbf{j}(\mu; Y_1, Y_2, Y_3)] = \mathbb{E}[3] = 3$$

Eksempel 2.2.2 Likelihoodfunktion for normalfordeling med ukendt varians Betragt den samme situation som ovenfor i eksempel 2.2.1, men antag nu, at der er udtaget n kuglelejer, og at variansen σ^2 er ukendt og skal estimeres fra data.

Log-likelihood funktionen er nu (idet der ses bort fra konstanten $-n \log(\sqrt{2\pi})$)

$$\ell(\mu, \sigma^2; y_1, y_2, \dots, y_n) = -\frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_i (y_i - \mu)^2$$

og idet vi benytter, at

$$\sum_i (y_i - \mu)^2 = n(\bar{y} - \mu)^2 + \sum_i (y_i - \bar{y})^2$$

får vi

$$\ell(\mu, \sigma^2; y_1, y_2, \dots, y_n) = -\frac{n}{2} \log(\sigma^2) - \frac{n}{2\sigma^2} (\bar{y} - \mu)^2 - \frac{1}{2\sigma^2} \sum_i (y_i - \bar{y})^2$$

der viser, at likelihoodfunktionen alene afhænger af de *sufficiente størrelser* $\bar{Y}$ og $\sum_i (Y_i - \bar{Y})^2$ med $\bar{Y} \sim N(\mu, \sigma^2/n)$ og $\sum_i (Y_i - \bar{Y})^2 \sim \sigma^2 \chi_{n-1}^2$ og hvor $\bar{Y}$ og $\sum_i (Y_i - \bar{Y})^2$ er uafhængige.

Scorefunktionen er

$$\ell' = \left\{ \begin{array}{c} \ell'_\mu(\mu, \sigma^2) \\ \ell'_{\sigma^2}(\mu, \sigma^2) \end{array} \right\} = \left\{ \begin{array}{c} \frac{n}{\sigma^2} (\bar{y} - \mu) \\ -\frac{n}{2\sigma^2} + \frac{n(\bar{y} - \mu)^2}{2\sigma^4} + \frac{\sum_i (y_i - \bar{y})^2}{2\sigma^4} \end{array} \right\}$$

hvorfor maksimum likelihood estimatet for (μ, σ^2) fås som $\hat{\mu} = \bar{y}$, og $\hat{\sigma}^2 = \sum_i (y_i - \bar{y})^2 / n$.

Men det følger af fordelingen for $\sum_i (Y_i - \bar{Y})^2$ at

$$\mathbb{E}[\hat{\sigma}^2] = \frac{n-1}{n} \sigma^2$$

Estimatoren er altså ikke *central*, men foretager en systematisk underestimation.

Ved estimationen af σ^2 kan man opfatte μ som en “nuisance” parameter uden relevans for σ^2 . Profillikelihood’en for variansen σ^2 er

$$\ell^*(\sigma^2) = -\frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_i (y_i - \bar{y})^2 \quad (2.16)$$

der fremkommer af likelihoodfunktionen for σ^2 ved at indsætte $\bar{y}$ i stedet for μ . Profillikelihood’en svarer altså til likelihood’en hvis man kendte værdien af μ og vidste at den netop havde værdien givet ved $\bar{y}$. Ved betragtning af profillikelihood’en indregnes altså ikke noget tab ved at man ikke kender μ .

Betragter man scorefunktionen svarende til (2.16)

$$\frac{d}{d\sigma^2} \ell^*(\sigma^2) = -\frac{n}{2\sigma^2} + \frac{\sum_i (y_i - \bar{y})^2}{2\sigma^4}$$

ses det, at den har forventningsværdien $-1/(2\sigma^2) \neq 0$, hvilket ikke harmonerer med relationen (2.6), der gælder for en “rigtig” scorefunktion.

Man kan endvidere vise, at variansen af denne scorefunktion er

$$\mathbb{V} \left[\frac{d}{d\sigma^2} \ell^*(\sigma^2) \right] = \frac{n-1}{2\sigma^4}$$

der er mere optimistisk end den forventede information (2.8), som er $(n-2)/(2\sigma^4)$

Der er forskellige måder hvorpå man kan håndtere problemerne omkring nuisance parametre, se feks. Pawitan [24]. Her vil vi blot anføre at i den situation, vi her betragter, kan man betragte *den marginale likelihood*, nemlig likelihoodfunktionen

$$\ell_{\text{marg}}(\sigma^2) = -\frac{n-1}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_i (y_i - \bar{y})^2$$

svarende til $\sigma^2\chi^2$ -fordelingen af $\sum_i(Y_i - \bar{Y})^2$. Benyttes denne marginale likelihood-funktion finder man det sædvanlige variansestimat

$$\hat{\sigma}^2 = \frac{\sum_i (y_i - \bar{y})^2}{n - 1}$$

▽

Eksempel 2.2.3 Likelihoodfunktion for binomialfordeling I en kundeundersøgelse har man udspurgt 10 tilfældigt udvalgte kunder om deres tilfredshed med den ydede service. Blandt disse 10 kunder var 3 kunder utilfredse.

Man formulerer en **model**

- **Model, data:** $Y \sim \text{bin}(10, p)$

Frekvensfunktionen for binomialfordelingen er

$$f_Y(y; p) = P[Y = y] = \binom{10}{y} p^y (1 - p)^{10-y} \quad \text{for } y = 0, 1, 2, \dots, 10$$

hvorfor likelihoodfunktionen svarende til $y = 3$ er

$$L(p; y = 3) = \binom{10}{3} p^3 (1 - p)^7$$

og log-likelihood funktionen er

$$l(p; y = 3) = 3 \ln(p) + 7 \ln(1 - p) + \ln \left(\binom{10}{y} \right)$$

med scorefunktionen

$$l'_p(p; y = 3) = \frac{3}{p} - \frac{7}{1 - p} = \frac{3 - 10p}{p(1 - p)}$$

og den observerede information

$$j(p; y = 3) = \frac{-3}{p^2} - \frac{7}{(1 - p)^2} = \frac{3 - 10p}{p(1 - p)}$$

Og i bogstaver:

$$l(p; Y) = Y \ln(p) + (10 - Y) \ln(1 - p) + \ln \left(\binom{10}{Y} \right)$$

med scorefunktionen

$$l'_p(p; Y) = \frac{Y}{p} - \frac{10 - Y}{1 - p} = \frac{Y - 10p}{p(1 - p)}$$

og den observerede information

$$j(p; Y) = \frac{-Y}{p^2} - \frac{10 - Y}{(1 - p)^2}$$

hvilket fører til den forventede information

$$\mathbf{i}(p) = \mathbb{E} \left[\frac{-Y}{p^2} - \frac{10 - Y}{(1 - p)^2} \right] = \frac{-10p}{p^2} - \frac{10(1 - p)}{(1 - p)^2} = -\frac{10}{p(1 - p)}$$

idet $\mathbb{E}[Y] = 10p$.

▽

Kapitel 3

Lidt om den generelle lineære model (normalfordelingsteori)

3.0 Introduktion og oversigt

De simple normalfordelingsmodeller (regressionsanalyse- og variansanalysemodeller), man møder i indledende statistikkurser, kan sammenfattes i en omfattende modelklasse, *den generelle lineære model* (GLM). I dette kapitel skitseres de vigtigste resultater for denne modelklasse. Først introduceres dog den flerdimensionale normalfordelingsmodel i afsnit 3.1.

En af de vigtige egenskaber for modellerne er den simple geometriske beskrivelse og fortolkning af reduktionen fra én model $H_1 : \boldsymbol{\mu} \in L_1$ til en simplere model $H_2 : \boldsymbol{\mu} \in L_2$, hvor L_1 og L_2 er lineære underrum med $L_2 \subset L_1$. I afsnit 3.3 gives indledningsvist en beskrivelse af estimation som projektion på det underrum, der er udspændt af modellen, og opspaltningen af kvadratafgivelsessummen introduceres. I afsnit 3.4 beskrives fordelingsegenskaberne for estimerne, og det bemærkes specielt, at residualerne er uafhængige af parameterestimerne (og de fittede værdier).

Det sædvanlige F-test for modelreduktion udledes som likelihood-ratio testet for modelreduktion. I afsnit 3.6.4 beskrives ideen bag de såkaldte type III opspaltninger, og i afsnit 3.6.5 beskrives modelreduktion ved successive projektioner og tilsvarende type I opspaltninger af kvadratafgivelsessummer.

Mens den abstrakte formulering ved lineære underrum er praktisk til forståelse af modellernes fundamentale egenskaber, benyttes i praksis sædvanligvis en mere parametrisk fremstilling, fx ved brug af modelformler. I afsnit 3.7 beskrives til referenceformål forskellige repræsentationer og metoder til parametrisering af modellerne. En parametriseret model refererer til en modelmatrix; en modelformel er lidt mere koordinatfri (den genererer sin egen modelmatrix)

Det er imidlertid vigtigt at holde sig for øje, at en given model (et givet underrum) kan parametriseres på mange måder, men de fittede værdier afhænger ikke af parametriseringen, men kun af underrummet.

Eksemplet i afsnit 3.8 kan benyttes også ved gennemlæsning af de foregående abstrakte afsnit til at konkretisere begreberne.

Endelig giver afsnit 3.9 lidt forklaring til nogle af de mange diagnostiske størrelser, der benyttes ved tilpasning af modellerne.

Vi bemærker at under normalfordelingsantagelsen gælder de udledte fordelingsresultater (F-test mv.) eksakt, dvs også for små stikprøver.

3.1 Statistisk model, den flerdimensionale normalfordelingsmodel

Ofte skrives en normalfordelingsmodel for k observationer $y_1, \dots, y_k$ på matrixform:

$$\mathbf{Y}_{k,1} = \boldsymbol{\mu}_{k,1} + \boldsymbol{\epsilon}_{k,1}$$

(Denne første gang har vi skrevet dimensioner på matricer og vektorer).

I mange anvendelser antages Y_i at være uafhængige, og med samme varians, σ^2 , dvs

$$\boldsymbol{\epsilon} \sim N_k(0_{k,1}, \sigma^2 \mathbf{I}_{k,k}) \quad (3.1)$$

I den generelle form tillades afhængighed mellem Y 'erne udtrykt ved en varians-kovariansmatrix, $\sigma^2 \boldsymbol{\Sigma}$, som er kendt på nær en faktor, σ^2 , dvs

$$\boldsymbol{\epsilon} \sim N_k(\mathbf{0}_{k,1}, \sigma^2 \boldsymbol{\Sigma}_{k,k}) , \quad (3.2)$$

hvor Σ er positiv definit.

Denne generelle form optræder for eksempel ved modellering af resultater af *blokforsøg*, eller andre situationer med hierarkisk (tilfældig) variation.

Af hensyn til brugen i de generaliserede lineære modeller vil vi hellere se på strukturen i middelværdierne i stedet for at sige observation = middelværdi PLUS støj:

$$\mathbf{Y} \sim N_k(\boldsymbol{\mu}, \sigma^2 \Sigma) \quad (3.3)$$

Definition 3.1.1 Tæthed for normalfordeling

Ved en normalfordelingsmodel, $\mathbf{Y} \sim N_k(\boldsymbol{\mu}, \sigma^2 \Sigma)$ for observationssættet $Y_1, Y_2, \dots, Y_n$, vil vi forstå en statistisk model, hvor den simultane tæthed for observations-sættet er af formen

$$f(\mathbf{y}; \boldsymbol{\mu}, \sigma^2) = \frac{1}{(\sqrt{2\pi})^k \sigma^k \det(\Sigma)} \exp \left[\frac{-1}{2\sigma^2} (\mathbf{y} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu}) \right] \quad \text{for } \mathbf{y} \in \mathbb{R}^k \quad (3.4)$$

▽

Varians- kovariansmatricen for observationssættet er

$$D[\mathbf{Y}] = \sigma^2 \Sigma$$

Den bilineare form

$$\delta_\Sigma(\mathbf{y}_1, \mathbf{y}_2) = \mathbf{y}_1' \Sigma^{-1} \mathbf{y}_2 \quad (3.5)$$

definerer et *indre produkt* på $\mathbb{R}^k$. Til dette indre produkt svarer et *ortogonalitetsbegreb* (svarende til at det indre produkt er nul), og en *norm* (defineret ved $\|\mathbf{y}\|_\Sigma = \sqrt{\delta(\mathbf{y}, \mathbf{y})}$).

Vi indfører notationen

$$D(\mathbf{y}; \boldsymbol{\mu}) = \|\mathbf{y} - \boldsymbol{\mu}\|_\Sigma^2 = (\mathbf{y} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu}) \quad (3.6)$$

til at symbolisere den kvadratiske norm af *vektoren* $(\mathbf{y} - \boldsymbol{\mu})$ svarende til det

indre produkt defineret ved Σ^{-1} .

Med disse betegnelser kan den simultane tæthed (3.4) udtrykkes som en tæthed på et vilkårligt endeligdimensionalt vektorrum for hvilket der er defineret et indre produkt, $\delta(\cdot, \cdot)$

$$f(\mathbf{y}; \boldsymbol{\mu}, \sigma^2) = \frac{1}{(\sqrt{2\pi})^k \sigma^k \det(\Sigma)} \exp \left[\frac{-1}{2\sigma^2} D(\mathbf{y}; \boldsymbol{\mu}) \right] \quad (3.7)$$

Bemærkning 1 *Scorefunktion og informationsmatrix for middelværdiparametrene*

Logaritmen til likelihoodfunktionen bliver

$$\ell_{\mu}(\boldsymbol{\mu}, \sigma^2; \mathbf{y}) = -(k/2) \log(\sigma^2) - \frac{1}{2\sigma^2} (\mathbf{y} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu}) \quad (3.8)$$

Scorefunktionen mht $\boldsymbol{\mu}$ er derfor

$$\frac{\partial}{\partial \boldsymbol{\mu}} \ell_{\mu}(\boldsymbol{\mu}, \sigma^2; \mathbf{y}) = \frac{1}{\sigma^2} [\Sigma^{-1} \mathbf{y} - \Sigma^{-1} \boldsymbol{\mu}] = \frac{1}{\sigma^2} \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu}) \quad (3.9)$$

og den observerede information (m.h.t. $\boldsymbol{\mu}$) er

$$\mathbf{j}(\boldsymbol{\mu}; \mathbf{y}) = \frac{1}{\sigma^2} \Sigma^{-1} . \quad (3.10)$$

Den observerede information (likelihoodfunktionens krumning) afhænger således ikke af observationerne $\mathbf{y}$, hvorfor den forventede information fås som

$$\mathbf{i}(\boldsymbol{\mu}) = \frac{1}{\sigma^2} \Sigma^{-1} .$$

▽

Det fremgår i øvrigt af (3.9), at under modellen (3.3) fås maksimum-likelihood estimatet for sættet af middelværdier ved at sætte observationerne lig med deres middelværdier.

3.2 Den generelle lineære model

Definition 3.2.1 *Generel lineær model, GLM*

En generel lineær model for $Y_1, Y_2, \dots, Y_k$ er en normalfordelingsmodel af formen (3.3), hvor der er formuleret en affin hypotese vedrørende middelværdierne $\mu_1, \mu_2, \dots, \mu_k$. Hypotesen er af formen

$$H_0 : \boldsymbol{\mu} - \boldsymbol{\mu}_0 \in L, \quad (3.11)$$

hvor L er et lineært underrum af $\mathbb{R}^k$ af dimension m , og hvor $\boldsymbol{\mu}_0$ angiver en vektor af kendte offsetværdier. ∇

Definition 3.2.2 *Dimension af generel lineær model*

Betragt den generelle lineære model givet ved (3.11). Dimensionen m af underrummet L kaldes for modellens dimension. ∇

Vi har valgt at formulere modellen i abstrakt vektorrumsterminologi, for så kan man tænke i projektioner og indre produkter m.v. uden at behøve at tænke på, hvordan man har valgt basis i L , og man kan formulere delhypoteser som under- rum $L_1 \subset L$, etc.

En sådan abstrakt angrebsvinkel er specielt nyttig fordi vi (og vores statistikprogram) ofte vælger at *overparametrisere* en model. Tænk bare på et faktorforsøg, eller envejsklassificerede data, hvor modellen ofte parametriseres som

$$E[Y_{ij}] = \mu + \alpha_i$$

og vi derfor bliver nødt til at indføre et (kunstigt) lineært bånd, fx $\sum n_i \alpha_i = 0$, eller $\alpha_k = 0$ for at sikre entydighed af modellen.

Definition 3.2.3 *Modelmatrix for generel lineær model*

Betragt en generel lineær model med en hypotese af formen (3.11).

Såfremt der er valgt en basis for L , sådan at hypotesen kan udtrykkes på formen

$$H_0 : \boldsymbol{\mu} - \boldsymbol{\mu}_0 = \mathbf{X}\boldsymbol{\beta} \quad \text{med } \boldsymbol{\beta} \in \mathbb{R}^m, \quad (3.12)$$

hvor $\mathbf{X}$ har fuld rang, kaldes matricen $\mathbf{X}$ for hypotesens modelmatrix.

Vektoren

$$\mathbf{x}_i^* = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{im} \end{pmatrix}, \quad (3.13)$$

hvis elementer er den i 'te række i modelmatricen, kaldes modelvektoren for den i 'te observation. ∇

Bemærkning 1 *Betydningen af offsetværdien μ_0*

Ved mange sædvanlige hypoteser er offset værdien μ_0 lig med nulvektoren $\mathbf{0}$, og hypotesen udtrykker, at prædiktoren μ ligger i et m -dimensionalt *under- rum* af $\mathbb{R}^k$, udspændt af søjlerne i $\mathbf{X}$.

Hvis offsetværdien μ_0 er forskellig fra nulvektoren, specificerer hypotesen, at prædiktoren μ ligger i et m dimensionalt *sideunderrum* i $\mathbb{R}^k$. ∇

I det følgende vil vi se bort fra offset-værdien, μ_0 , idet vi for normalfordelingsmodellen blot kan foretage en nulpunktsforskydning af Y 'erne og derved opnå $\mu_0 = \mathbf{0}$.

Bemærkning 2 *Formulering af hypotese ved lineære bånd.*

Vi bemærker, at en hypotese af formen (3.12) alternativt kunne specificeres ved

$$\mathbf{M}\mu = \mathbf{M}\mu_0 \quad (3.14)$$

hvor $\mathbf{M}$ er en $(k - m) \times m$ -dimensional matrix af fuld rang.

Rummet udspændt af (3.12) er jo netop løsningsmængden svarende til et ligningssystem af formen (3.14). ∇

Bemærkning 3 *Modelmatricen behøver ikke have fuld rang* Det er kun af bekvemmelighedsgrunde, at ovenstående definition antager, at $\mathbf{X}$ har fuld rang. Det afgørende er, at søjlerne i $\mathbf{X}$ er basisvektorer for L . Når vi har valgt at antage fuld rang er det for at kunne tillade sig at skrive $(\mathbf{X}'\mathbf{X})^{-1}$ uden at skulle bøvl med generaliserede inverser. ∇

Bemærkning 4 *Geometrisk fortolkning*

Observationssættet $y_1, \dots, y_k$ kan opfattes som et punkt i det k -dimensionale euklidiske rum $\mathbb{R}^k$ (observationsrummet).

Matricen $\mathbf{X}$ udspænder et lineært underrum, $L \subset \mathbb{R}^k$. (For nemheds skyld kan man forestille sig, at $\mathbf{X}$ har fuld rang i hvilket tilfælde underrummet L har dimensionen m).

Når parameteren β gennemløber parameter rummet $\beta \in \mathbb{R}^m$, da vil de tilsvarende værdier af middelværdien $\mu = \mathbf{X}\beta$ beskrive en m -dimensional hyperplan, $L = \mu(B) \subset \mathbb{R}^k$. Hyperplanen er udspændt af søjlerne i $\mathbf{X}$, og β angiver de *lokale koordinater* i det koordinatsystem, der er angivet ved søjlerne i $\mathbf{X}$.

▽

Fuld model og minimal model

To yderpunkter i samlingen af potentielle modeller er den *fulde model*, der siger at enhver observation har sin egen middelværdi, og den *minimale model*, der siger at alle observationerne har samme middelværdi (vi ser normalt bort fra den endnu mindre model, hvor alle observationer har middelværdien nul)

Definition 3.2.4 *Fuld model* Ved den fulde model vil vi forstå modellen svarende til at middelværdivektoren μ varierer frit i $\mathbb{R}^k$, dvs modellen med modelmatricen

$$\mathbf{X} = \mathbf{I}_k$$

▽

Definition 3.2.5 *Minimal model* Ved den minimale model vil vi forstå modellen svarende til den udartede modelmatrix

$$\mathbf{X} = \begin{pmatrix} 1 \\ 1 \\ \dots \\ 1 \end{pmatrix}$$

▽

For $\Sigma = \mathbf{I}$ er afstanden fra observationspunktet til den minimale model netop kvadratafgivelsessummen omkring det fælles gennemsnit.

3.3 Estimation i en GLM

Under hypotesen (3.11)

$$H_0 : \boldsymbol{\mu} \in L ,$$

bestemmes maksimum-likelihood estimatet for sættet af middelværdier som den ortogonale (m.h.t δ_Σ) projektion, $p_L(\mathbf{y})$ af punktet $\mathbf{y}$ ned på det lineære underrum, L .

Såfremt hypotesen er udtrykt på formen (3.12)

$$H_0 : \boldsymbol{\mu}(\boldsymbol{\beta}) = \mathbf{X}\boldsymbol{\beta}$$

bestemmes maksimum-likelihood estimatet for $\boldsymbol{\beta}$ som løsning til *normalligningen*

$$\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y} = \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}} \quad (3.15)$$

Altså m ligninger med m ubekendte.

Hvis $\mathbf{X}$ har fuld rang (m), er der en entydig løsning, som bestemmes ved

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y} \quad (3.16)$$

Bevis:

Det følger umiddelbart af udtrykket (3.7) for likelihoodfunktionen, at maksimum mht $\boldsymbol{\mu}$ fås ved at minimere "afvigelsen" $D(\mathbf{y}; \boldsymbol{\mu})$ for $\boldsymbol{\mu} \in L$. Da $D(\mathbf{y}; \boldsymbol{\mu})$ udtrykker en norm mht til det indre produkt, δ_Σ , opnås minimum netop ved at projicere ned på L .

For en parametriseret hypotese (3.12) finder man ved benyttelse af (2.11), idet

$$\frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{\mu}(\boldsymbol{\beta}) = \frac{\partial}{\partial \boldsymbol{\beta}} \mathbf{X}\boldsymbol{\beta} = \mathbf{X}$$

at

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{\beta}} \ell_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \sigma^2; \mathbf{y}) &= \left[\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}} \right]' \frac{\partial}{\partial \boldsymbol{\mu}} \ell_{\boldsymbol{\mu}}(\boldsymbol{\mu}(\boldsymbol{\beta}), \sigma^2; \mathbf{y}) \\ &= \frac{1}{\sigma^2} \mathbf{X}' [\boldsymbol{\Sigma}^{-1} \mathbf{y} - \boldsymbol{\Sigma}^{-1} \mathbf{X} \boldsymbol{\beta}]\end{aligned}\quad (3.17)$$

idet scorefunktionen mht $\boldsymbol{\mu}$ er givet ved (3.9).

Det ses, at (3.17) netop er nul, når normalligningen (3.15) er opfyldt. ∇

Bemærkning 1 *Fittede værdier, hatmatrix*

De *fittede værdier* $\hat{\boldsymbol{\mu}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ fås ved at projicere $\mathbf{y}$ ned på planen udspændt af $\mathbf{X}$, og $\hat{\boldsymbol{\beta}}$ angiver de lokale koordinater (i koordinatsystemet bestemt af søjlerne i $\mathbf{X}$ matricen) til projektionen.

I SAS-Insight fremkommer de fittede værdier i outputvariablen `Predicted`.

Hatmatricen

$$\mathbf{H} = \mathbf{X}[\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}]^{-1}\mathbf{X}'$$

er projektionsmatricen, der overfører $\mathbf{y}$ til de fittede værdier $\hat{\boldsymbol{\mu}}$ (nemlig sætter hat på y 'erne).

idet nemlig

$$\hat{\boldsymbol{\mu}} = p_L(\mathbf{y}) = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{H}\mathbf{y}$$

I det simple tilfælde med $\boldsymbol{\Sigma} = \mathbf{I}$ er hatmatricen blot

$$\mathbf{H} = \mathbf{X}[\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}'$$

I SAS-Insight fremkommer diagonaleelementerne i hatmatricen i outputvariablen `Hat diag`.

Residualerne er

$$\mathbf{r} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} = (\mathbf{I} - \mathbf{H})\mathbf{y} \quad (3.18)$$

I SAS-Insight fremkommer residualerne i outputvariablen **Residual**.

Hat-matricen er en projektionsmatrix og derfor er den symmetrisk (nemlig $\mathbf{H}' = \mathbf{H}$) og idempotent ($\mathbf{H}^2 = \mathbf{H}$)

▽

Bemærkning 2 *Ortogonalitet*

Maksimum-likelihood estimatet for β bestemmes åbenbart som den værdi af β , der minimerer *afstanden* $\|\mathbf{y} - \mathbf{X}\beta\|$ under denne norm.

Det følger af normalligningen (3.15), at

$$\mathbf{X}'\Sigma^{-1}(\mathbf{y} - \mathbf{X}\hat{\beta}) = \mathbf{0} \quad (3.19)$$

dvs, at *residualerne* (3.18) er *ortogonale* (mht Σ^{-1}) på planen, L , udspændt af $\mathbf{X}$.

Dvs specielt er residualvektoren ortogonal på vektoren af fittede værdier. ▽

Bemærkning 3 *Dimension (rang) af residualvektor*

Residualerne $\mathbf{r}$ ligger altså i et lineært underrum af dimensionen $k - m$, som er ortogonalt på L (af dimension m).

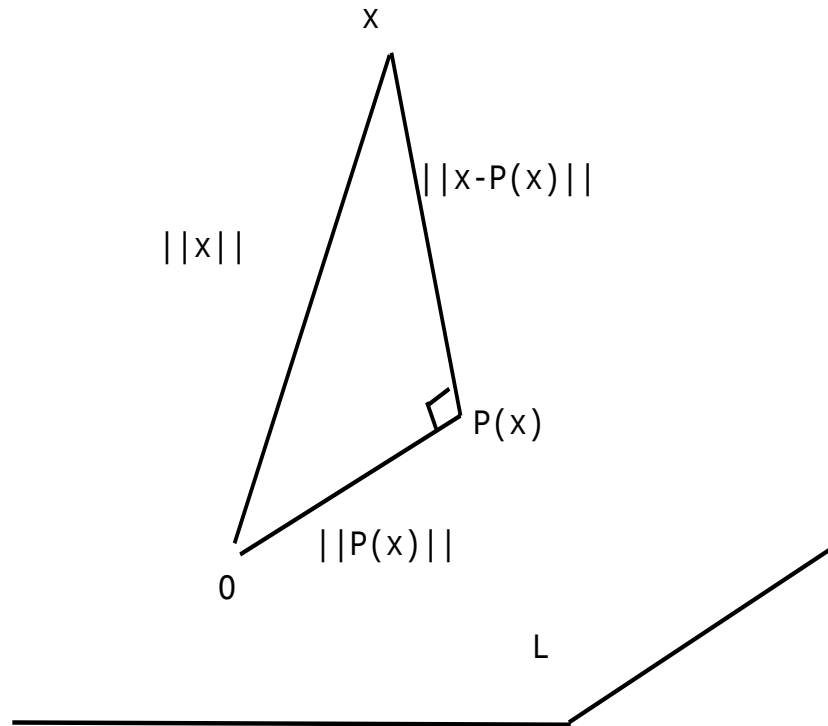
▽

Bemærkning 4 *Opspaltning af kvadratsum*

Det følger af ovenstående betragtninger, at der gælder den pythagoræiske opspaltning

$$\begin{aligned} D(\mathbf{y}; \mathbf{0}) &= D(\mathbf{y}; \mathbf{X}\hat{\beta}) + D(\mathbf{X}\hat{\beta}; \mathbf{0}) \\ &= (\mathbf{y} - \mathbf{X}\hat{\beta})'\Sigma^{-1}(\mathbf{y} - \mathbf{X}\hat{\beta}) + (\mathbf{X}\hat{\beta})'\Sigma^{-1}\mathbf{X}\hat{\beta} \end{aligned} \quad (3.20)$$

af den totale variation omkring nul i en residualvariation og en variation af estimerterne omkring nul som illustreret i nedenstående figur



De sædvanlige variansanalysekemaer beskriver en tilsvarende opspaltning af variationen omkring den minimale model (det fælles gennemsnit).

▽

3.4 Fordeling af $\hat{\beta}$, fittede værdier og residualer, estimation af σ^2

3.4.1 Fordeling af $\hat{\beta}$, fittede værdier og residualer

Da $\hat{\beta}$ er en linearkombination af de normalt fordelte Y 'er gælder at fordelingen af linearkombinationen $\hat{\beta}$ af Y 'erne er en $N_m(\beta, \sigma^2 (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1})$ -fordeling, dvs middelværdi og dispersionsmatrix for $\hat{\beta}$ er givet ved

$$E[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta} \quad (3.21)$$

$$D[\hat{\boldsymbol{\beta}}] = \sigma^2 (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1} \quad (3.22)$$

Og ligeledes gælder, at de fittede værdier er normalt fordelt med

$$D[\mathbf{X}\hat{\boldsymbol{\beta}}] = \sigma^2 \mathbf{X}[\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}]^{-1}\mathbf{X}' = \sigma^2 \mathbf{H}$$

der har rangen m .

Og residualerne $\mathbf{r} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ er uafhængige af de fittede værdier og har middelværdien nul og dispersionsmatricen

$$D[\mathbf{r}] = \sigma^2 (\mathbf{I} - \mathbf{H})$$

med rangen $k - m$.

Residualerne har således ikke samme varians. Det i 'te residual har variansen

$$V[r_i] = \sigma^2 (1 - h_{ii}) \quad (3.23)$$

hvor h_{ii} er det i 'te diagonalelement i hat-matricen. Hvis specielt h_{ii} er nær ved 1, er der altså ikke stor varians på residuallet.

Modsvarende har vi, at

$$V[\hat{\mu}_i] = \sigma^2 h_{ii} ;$$

dvs, at hvis h_{ii} er lille, er der ikke stor varians på den fittede værdi.

Vi finder specielt, at *korrelationen mellem en observation og den fittede værdi* er

$$\rho(y_i, \hat{\mu}_i) = \sqrt{h_{ii}}$$

3.4.2 Kvadratiske former, spaltningsætningen

For opspaltningen af kvadratafvigelsessummen (3.20) gælder den såkaldte *spaltningsætning for kvadratiske former*, se fx Christensen [1] Afsnit 1.3.

Essensen i sætningen er, at for $\mathbf{Y} \sim N_k(\mathbf{0}, \boldsymbol{\Sigma})$, og for det tilsvarende indre

produkt, $\delta(\cdot, \cdot)$, og norm defineret som før ved (3.5) og $\|\mathbf{y}\|_{\Sigma} = \sqrt{\delta(\mathbf{y}, \mathbf{y})}$, vil den kvadratiske form

$$\|\mathbf{Y}\|_{\Sigma}^2 = \mathbf{Y}'\Sigma^{-1}\mathbf{Y} \sim \chi^2(k)$$

og yderligere, hvis $\mathbf{R}^k$ er en direkte sum af de ortogonale (m.h.t $\delta(\cdot, \cdot)$) underrum $U_1, \dots, U_m$, da vil normen Q_i af projektionen af $\mathbf{Y}$ ned på U_i være $\chi^2(\dim(U_i))$ fordelt, og yderligere vil normerne Q_i og Q_j være uafhængige (for $i \neq j$).

Hvis altså vi har *spaltet* observationsrummet i en direkte sum af en række ortogonale underrum, kan vi (som allerede Pythagoras kunne) spalte den χ^2 -fordelte afstand $\|\mathbf{y}\|_{\Sigma}^2$ i en tilsvarende sum af uafhængige χ^2 -fordelte størrelser, hvis frihedsgrader er dimensionerne af de pågældende underrum.

3.4.3 Fordeling af residualkvadratsummen, estimation af σ^2

Det følger nu af (3.20) at residualkvadratsummen følger en $\sigma^2\chi^2(k - m)$ -fordeling

$$D(\mathbf{y}; \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{r}'\Sigma^{-1}\mathbf{r} \sim \sigma^2\chi^2(k - m)$$

og at residualkvadratsummen er *uafhængig* af estimaterne for $\boldsymbol{\beta}$.

Ved simultan maksimum-likelihood estimation af $\boldsymbol{\beta}$ og σ^2 finder man maksimum-likelihood estimatoren

$$\hat{\sigma}^2 = \mathbf{r}'\Sigma^{-1}\mathbf{r}/k = D(\mathbf{y}; \mathbf{X}\hat{\boldsymbol{\beta}})/k$$

og den har de sædvanlige gode asymptotiske egenskaber for maksimum-likelihood estimatorer, men alligevel er den utilfredsstillende, hvis antallet, m , af β -parametre ikke er forsvindende i forhold til antallet, k , af observationer. Blandt andet er den kun asymptotisk central (dvs. med den rette middelværdi).

I en sådan situation med *nuisance parametre* $\boldsymbol{\beta}$, der er irrelevante for estimationen af σ^2 , vil man tage hensyn til at de er maksimeret væk i profil-likelihood'en, og derfor vil man ikke direkte bruge profillikelihood'en, men

benytte at fordelingen af residualerne er uafhængig af fordelingen af de fittede værdier og af nuisanceparametrene β , og så kan man bare kigge på den *marginale likelihood* svarende til $\sigma^2\chi^2$ -fordelingen af residualkvadratsummen i analogi med eksempel 2.2.2.

Man benytter derfor

$$s^2 = \mathbf{r}'\Sigma^{-1}\mathbf{r}/(k - m) = D(\mathbf{y}; \mathbf{X}\hat{\beta})/(k - m) \quad (3.24)$$

som estimator for σ^2 .

3.4.4 Test af enkelte parametre

Det følger af ovenstående resultater om fordelingen af estimatorne, at hypoteser vedrørende enkelte parametre β_j kan testes ved t-test i de marginale normalfordelinger af $\hat{\beta}_j$. Sædvanligvis angiver SAS-outputtet under overskriften “parameter estimates” teststørrelser (t-test) og p -værdier for sådanne test for hypoteser af formen $\beta_j = 0$

3.5 Eksempel, simpel lineær regressionsmodel

Højde af ramusknogle

Nedenstående tabel viser målinger af højden af ramusknoglen hos en dreng. Målingerne er foretaget med halvårige intervaller fra hans 8-års fødselsdag.

Alder [År]	højde [mm]
8	51.2
8 1/2	53.0
9	54.3
9 1/2	54.5

Vi prøver nu som et regneeksempel at udføre en lineær regressionsanalyse af ramushøjde mod alder. (Dette er dog ikke en særlig hensigtsmæssig model for vækst plus målefejl)

Antager vi nu, at ramushøjden Y_i ved alderen x_i^* kan beskrives ved en $N(\mu_i, \sigma^2)$ fordelt variabel, har vi modellen

$$\mu_i = \beta_1^* + \beta_2 x_i^*, i = 1, 2, \dots, 4$$

Vi vil illustrere brugen af modelmatricen til bestemmelse af løsningen.

For at simplificere beregningerne vælger vi at regne alderen $x_i = x_i^* - 8.75$ (dvs alderen i år minus 8.75). Denne transformation ændrer ikke analysen, men det er klart, at de fundne værdier af koefficienterne skal fortolkes i lys af dette valg af nulpunkt.

Med dette valg har vi modelmatricen:

$$\mathbf{X} = \begin{pmatrix} 1.0 & -0.75 \\ 1.0 & -0.25 \\ 1.0 & 0.25 \\ 1.0 & 0.75 \end{pmatrix} \quad \text{med} \quad \mathbf{X}'\mathbf{X} = \begin{pmatrix} 4.00 & 0.00 \\ 0.00 & 1.25 \end{pmatrix}$$

hvorfor

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} 0.25 & 0.00 \\ 0.00 & 0.80 \end{pmatrix},$$

der fører til estimerne $\hat{\beta}_1 = 53.25$, $\hat{\beta}_2 = 2.24$

De fittede værdier $\hat{\mu}_i$ bestemmes da som

$$\hat{\mu}_i = 53.25 + 2.24 x_i$$

og residualkvadratet svarende til den i 'te observation er

$$d(y_i, \hat{\mu}_i) = (y_i - 53.25 - 2.24 x_i)^2$$

De fittede værdier og residualkvadraterne er angivet i nedenstående tabel (residualkvadraterne kaldes her *devianser* i overensstemmelse med terminologien for generaliserede lineære modeller):

Alder [År]		højde [mm]	fittet værdi	deviansbidrag
	x_i	y_i	$\hat{\mu}_i$	$d(y_i, \hat{\mu}_i)$
8	-0.75	51.2	51.57	0.1369
8 1/2	-0.25	53.0	52.69	0.0961
9	0.25	54.3	53.81	0.2401
9 1/2	0.75	54.5	54.93	0.1849
sum				0.6580

Hvis vi kendte variansen, σ^2 , kunne vi vurdere modeltilpasningen ved at sammenholde residualkvadratsummen $\sum d(y_i; \hat{\mu}_i)$ med dens forventningsværdi $(4 - 2)\sigma^2$. Imidlertid kender vi ikke variansen, men såfremt vi antager at den lineære model kan opretholdes, kan vi estimere variansen σ^2 ved

$$\hat{\sigma}^2 = 0.6580 / (4 - 2) = 0.3290$$

svarende til

$$\hat{\sigma} = 0.5736 \text{ [mm]}$$

Modeltilpasningen kan vurderes ved at betragte residualerne. Således kan normalfordelingsantagelsen vurderes ved at tegne et fraktildiagram, et såkaldt QQ-plot, varianshomogeniteten kan vurderes ved at tegne residualerne mod de fittede værdier. I afsnit 3.9 angives en række metoder til kontrol af modellen.

Geometrisk

Observationsættet (de fire højder) kan opfattes som et punkt i det 4-dimensionale rum. Søjlerne i matricen $\mathbf{X}$ udspænder en plan i dette rum og danner basisvektorer for denne plan, hvor β_1 og β_2 angiver de lokale koordinater. Og fordi vi var så snedige at bruge alderen 8.75 som udgangspunkt er det lokale koordinatsystem ortogonalt.

Havde vi mon fået en anden plan, hvis vi havde regnet alderen i almindelige år ?

SAS-output

I SAS-Insight vælges Fit-menuen, og man formulerer modellen med ramushøjde som responsvariabel (Y) og den reducerede alder som forklarende variabel (X). Såfremt man i Output menuen havde markeret rubrikken **Analysis of Variance/Deviance**, havde man fået variansanalyseskemaet:

The SAS System

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Stat	Pr > F
Model	1	6.2720	6.2720	19.06	0.0487
Error	2	0.6580	0.3290	.	.
C Total	3	6.9300	.	.	.

der netop afspejler en opspaltning af den totale afvigelse omkring det fælles gennemsnit (**C Total**) i residualvariationen omkring den estimerede linie (**Error**) og liniens afvigelse fra vandret (**Model**), hvor frihedsgraderne (DF) netop angiver dimensionerne af de tilsvarende underrum.

En markering af **Parameter Estimates** giver udskrift af parameterestimerne, den estimerede spredning på disse estimater, en t-størrelse, der tester en hypotese om hvorvidt den pågældende parameter værdi kan antages at være nul, samt den tilsvarende signifikanssandsynlighed (*p*-værdi)

Parameter Estimates					Pr
Variable	DF	Estimate	Std Error	t Stat	> t
Intercept	1	53.2500	0.2868	185.67	<.0001
ald	1	2.2400	0.5130	4.37	0.0487

I menuen **Output variables** kan man blandt andet vælge at få beregnet de fittede værdier $\hat{\mu}_i$ og residualerne. Endvidere kan man vælge at få beregnet en række diagnostiske størrelser. Disse størrelser er beskrevet i afsnit 3.9.

3.6 Reduktion af en generel lineær model

3.6.1 Introduktion

Ved en vurdering af hvilke koefficienter, der er nødvendige til beskrivelse af interessevariablen, kan man naturligvis godt vurdere én koefficient ad gangen ved brug af de marginale t-størrelser, som anført i afsnit 3.4.4 ovenfor. Man skal dog være forsigtig med at bruge en sådan fremgangsmåde, da den jo ikke tilgodeser den korrelation, der er mellem de forskellige parameterestimer.

Endvidere skal man være meget forsigtig med at bruge metoden til at fjerne flere koefficienter på én gang. Den korrelation, der er mellem estimatorne af de enkelte koefficienter, kan bevirke, at selv om et antal af koefficienterne hver for sig kan antages at være nul, kan de ikke alle antages at være nul samtidig. En anden indvending, man kan rette mod denne fremgangsmåde er, at selv om det enkelte test udføres, fx. på et 5 % niveau, vil niveauet svarende til den enkeltvise vurdering af et antal koefficienter være noget mindre end de formelle 5 %. (Hvis alle koefficienter virkelig var nul, ville sandsynligheden for at mindst én af r koefficienter blev dømt som værende forskellig fra nul ved r uafhængige tests af koefficienterne være $1 - (0.95)^r$, altså væsentligt større end 5 %).

En fremgangsmåde kan naturligvis være at *reparametrisere* problemet således at den resulterende modelmatrix er ortogonal. Herved opnår man uafhængige estimater af de enkelte parametre. Denne fremgangsmåde benyttes ofte i forsøgsplanlægningen, når det er muligt at tilrettelægge dataindsamlingen på en måde, der tillader meningsfuld ortogonal parametrisering. Brugen af ortogonale polynomier er et andet eksempel på en situation, hvor en sådan fremgangsmåde er mulig. Ligeledes er regression efter principale komponenter et eksempel på en ortogonal parametrisering.

En anden fremgangsmåde er at formulere en række modeller, hvor de tilsvarende underrum ligger inden i hinanden, og derefter foretage successive modelreduktioner ved betinget test i disse modeller. I det følgende vil denne fremgangsmåde blive beskrevet nærmere.

3.6.2 Kvotienttest for modelreduktion

Kvotienttest

Vi vil benytte det såkaldte kvotienttest (likelihood-ratio test, se fx. Shao [13] eller Bickel og Doksum [3]): Lad Y være en stokastisk variabel, hvis fordeling under en forelagt model kan beskrives ved en familie af tætheder, $\{f(y; \theta)\}_{\theta \in \Omega}$ og betragt en hypotese (delmodel) specificeret ved $\theta \in \Omega_1 \subset \Omega$, hvor $\dim(\Omega) = k$ og $\dim(\Omega_1) = m$.

Kvotienttestet for hypotesen $H_0 : \theta \in \Omega_1$ imod alternativet $H_0^a : \theta \in \Omega \setminus \Omega_1$

bygger på kvotientteststørrelsen

$$q(y) = \frac{\sup_{\theta \in \Omega_1} L(\theta; y)}{\sup_{\theta \in \Omega} L(\theta; y)}$$

Små værdier af q fortolkes som *evidens imod hypotesen*. Testet afviser således for små værdier af q .

Der gælder med sandsynligheden 1, at

$$0 \leq q(y) < 1$$

Under regularitetsbetingelser (i det væsentlige, at støtten for f ikke afhænger af parametrene) vil $-2 \ln q(y)$ approksimativt følge en $\chi^2(k - m)$ fordeling. (Se fx Shao, [13]).

Fortolkningen af testet er, at det sammenligner maksimumværdien af likelihoodfunktionen under modellen (givet ved Ω) med maksimumværdien under hypotesen H_0 . Hvis maksimumværdien under hypotesen er væsentligt mindre end under modellen, tages dette som udtryk for at observationerne ikke er i overensstemmelse med hypotesen. Altså, jo mindre værdi af q , desto mindre overensstemmelse mellem data og hypotese. Testets *niveau* fastlægger afskærringsværdien af q (eller af $-2 \log q$).

I den generelle lineære model kan man angive den eksakte fordeling for teststørrelsen, som vist nedenfor.

Maksimum af likelihoodfunktionen

Betragt normalfordelingsmodellen (3.4) $\mathbf{Y} \sim N_k(\boldsymbol{\mu}, \sigma^2 \boldsymbol{\Sigma})$ for den fulde model. Lad $\mathbf{y} \in \mathbb{R}^k$ og betragt et underrum $L \subset \mathbb{R}^k$. Vi vil lade $p_L(\mathbf{y})$ betegne den ortogonale (m.h.t dette indre produkt) *projektion* af $\mathbf{y}$ ned på L .

Hvis underrummet L er udspændt af modelmatricen $\mathbf{X}$, så vi i afsnit 3.3, at

$$p_L(\mathbf{y}) = \mathbf{H}\mathbf{y} = \mathbf{X}[\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}]^{-1}\mathbf{X}'\mathbf{y} \quad (3.25)$$

For $\Sigma = \mathbf{I}_k$ er det bare den sædvanlige projektion.

Såfremt L er parametriseret ved parametrene β svarende til modelmatricen $\mathbf{X}$ bliver logaritmen til likelihoodfunktionen

$$\ell(\beta, \sigma^2) = -(k/2) \log(\sigma^2) - \frac{1}{2\sigma^2} D(\mathbf{y}; \mu)$$

Vi har allerede fundet maksimum-likelihood estimatet for β og konstateret at estimationen af β ikke involverer σ^2 .

Maksimum-likelihood estimatet for σ^2 er

$$\hat{\sigma}^2 = \frac{1}{k} D(\mathbf{y}; p_L(\mathbf{y})) \quad (3.26)$$

idet jo de fittede værdier under modellen svarende til L er $\hat{\mu} = p_L(\mathbf{y})$.

Værdien af likelihoodfunktionen svarende til maksimumværdien fås ved indsættelse af $\hat{\mu} = p_L(\mathbf{y})$ og $\hat{\sigma}^2 = D(\mathbf{y}; p_L(\mathbf{y}))/k$. Det ses, at for maksimumspunktet bliver den faktor i likelihoodfunktionen, der indeholder eksponentialfunktionen, simpelthen bare $\exp(-k/2)$. Logaritmen til likelihoodfunktionen i maksimumspunktet bliver da

$$\ell(\hat{\beta}, \hat{\sigma}^2) = -(k/2) \log(\hat{\sigma}^2) - \frac{k}{2} \quad (3.27)$$

hvor $\hat{\sigma}^2$ er bestemt ved (3.26)

Vi bemærker, at da de fittede værdier kun afhænger af underrummet, og ikke af den valgte parametrisering (basis, modelmatrix), afhænger maksimumsværdien af likelihoodfunktionen heller ikke af den valgte parametrisering af underrummet.

Sætning 3.6.1 *Kvotienttest for modelreduktion i generel lineær model*

Betragt normalfordelingsmodellen $\mathbf{Y} \sim N_k(\mu, \sigma^2 \Sigma)$ og antag, at der gælder modellen

$$H_1 : \mu \in L \subset \mathbb{R}^k$$

med $\dim(L) = m_1$.

Lad $L_2 \subset L$ være et lineært (affint) underrum med $\dim(L_2) = m_2$. Da er kvotientteststørrelsen for hypotesen

$$H_2 : \boldsymbol{\mu} \in L_2 \quad \text{imod alternativet } H_2^a : \boldsymbol{\mu} \in L \setminus L_2$$

en monoton funktion af

$$F(\mathbf{y}) = \frac{D(p_1(\mathbf{y}); p_2(\mathbf{y})) / (m_1 - m_2)}{D(\mathbf{y}; p_1(\mathbf{y})) / (k - m_1)} \quad (3.28)$$

hvor $p_1(\mathbf{y})$ og $p_2(\mathbf{y})$ angiver projektionerne af $\mathbf{y}$ på henholdsvis L og L_2 .

Under hypotesen H_2 vil $F \sim F(m_1 - m_2, k - m_1)$; dvs store værdier af F er udtryk for manglende overensstemmelse mellem data og hypotesen.

Bevis:

Det følger af (3.27), at kvotientteststørrelsen er

$$q_{2|1}(\mathbf{y}) = \left[\frac{D(\mathbf{y}; p_1(\mathbf{y}))}{D(\mathbf{y}; p_2(\mathbf{y}))} \right]^{k/2}$$

men da $p_1(\mathbf{y}) \in L$, og $p_2(\mathbf{y}) \in L_2 \subset L$, vil også $p_1(\mathbf{y}) - p_2(\mathbf{y}) \in L$. Da endvidere residualvektoren $\mathbf{y} - p_1(\mathbf{y})$ er ortogonal på L , er $\mathbf{y} - p_1(\mathbf{y})$ og $p_1(\mathbf{y}) - p_2(\mathbf{y})$ ortogonale, og vi har derfor den Pythagoræiske relation,

$D(\mathbf{y}; p_2(\mathbf{y})) = D(\mathbf{y}; p_1(\mathbf{y})) + D(p_1(\mathbf{y}); p_2(\mathbf{y})) \quad (3.29)$
--

således at vi har

$$\begin{aligned} q_{2|1}(\mathbf{y}) &= \left[\frac{D(\mathbf{y}; p_2(\mathbf{y}))}{D(\mathbf{y}; p_1(\mathbf{y})) + D(p_1(\mathbf{y}); p_2(\mathbf{y}))} \right]^{k/2} \\ &= \left[\frac{1}{1 + \frac{D(p_1(\mathbf{y}); p_2(\mathbf{y}))}{D(\mathbf{y}; p_1(\mathbf{y}))}} \right]^{k/2} = \left[\frac{1}{1 + \frac{m_1 - m_2}{k - m_1} F} \right]^{k/2} \end{aligned}$$

hvor F er givet ved (3.28). Altså jo større værdi af F , desto mindre værdi af kvotientteststørrelsen q .

Det følger af spaltningssætningen i forbindelse med (3.29), at da $D(\mathbf{y}; p_2(\mathbf{y})) \sim \sigma^2 \chi^2(k - m_2)$, og da $\mathbf{y} - p_1(\mathbf{y})$ og $p_1(\mathbf{y}) - p_2(\mathbf{y})$ er ortogonale, at $D(\mathbf{y}; p_1(\mathbf{y})) \sim \sigma^2 \chi^2(k - m_1)$ og $D(p_1(\mathbf{y}); p_1(\mathbf{y})) \sim \sigma^2 \chi^2(m_1 - m_2)$ er uafhængige, hvorfor det følger, at $F \sim F(m_1 - m_2, k - m_1)$.

▽

Lad $\hat{\boldsymbol{\mu}}$ og $\hat{\boldsymbol{\mu}}$ angive de fittede værdier under henholdsvis H_1 og H_2 . Vi kan da udtrykke relationen (3.29) som

$$D(p_1(\mathbf{y}); p_2(\mathbf{y})) = D(\hat{\boldsymbol{\mu}}; \hat{\boldsymbol{\mu}}) = D(\mathbf{y}; \hat{\boldsymbol{\mu}}) - D(\mathbf{y}; \hat{\boldsymbol{\mu}}) \quad (3.30)$$

Testet måler afvigelsen, $D(\hat{\boldsymbol{\mu}}; \hat{\boldsymbol{\mu}})$ mellem de fittede værdier $\hat{\boldsymbol{\mu}}$ under modellen, og de fittede værdier $\hat{\boldsymbol{\mu}}$ under hypotesen i forhold til den naturlige usikkerhed under modellen, $D(\mathbf{y}; \hat{\boldsymbol{\mu}})$.

Det statistik-tekniske, der ligger bag, er, at på grund af normalfordelingen af observationerne, den lineære model- og hypotese, ortogonalitetssegenskaber ved projektioner samt spaltningssætningen lykkedes det at *spalte* residualafvigelsen under hypotesen, $D(\mathbf{y}; \hat{\boldsymbol{\mu}})$, i to uafhængige bidrag, ét, der måler tilpasningen, $D(\mathbf{y}; \hat{\boldsymbol{\mu}})$, til modellen H_1 , og ét, der måler tilpasningen, $D(\hat{\boldsymbol{\mu}}; \hat{\boldsymbol{\mu}})$, til hypotesen H_2 under antagelse af, at modellen H_1 er sand. Opspaltningen, der danner grundlag for opskrivning af *variansanalysekemaer* er illustreret i figur 3.1.

Testet består således i at vurdere forøgelsen i residualkvadratsum ved reduktion af modellen til H_2 i forhold til residualkvadratsummen under H_2 . Vi kan formulere testet i ord:

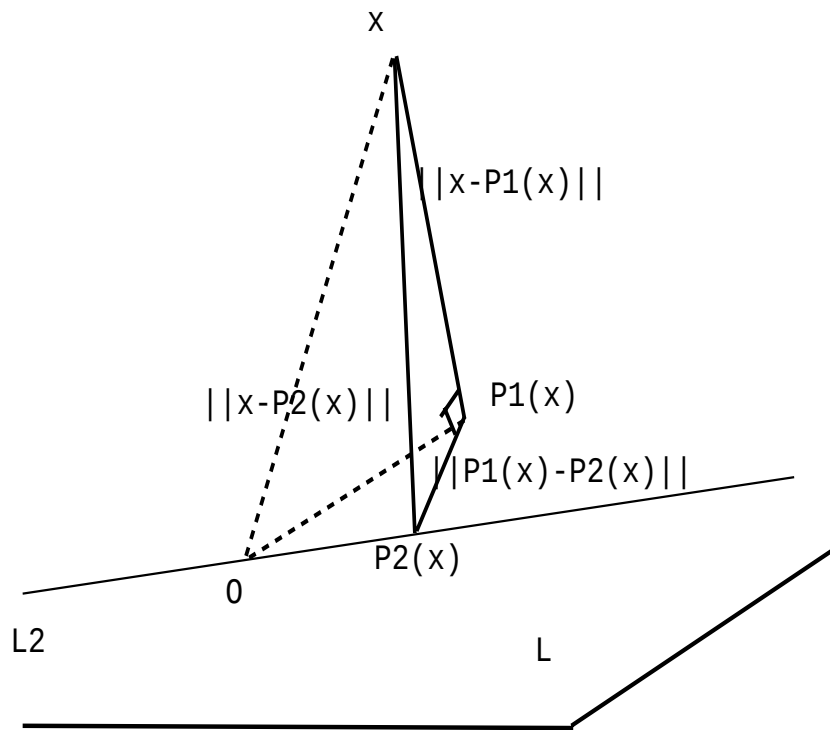
Test for udeladelse af led:

F-størrelse:

$$F = \frac{\text{forøgelse i residualkvadratsum/antal testede parametre}}{\text{estimat for } \sigma^2 \text{ fra model}}$$

Det skal bemærkes, at testet er udledt som et *betinget test*. Det tester hypotesen $H_2 : \boldsymbol{\mu} \in L_2$ under forudsætning af, at $H_1 : \boldsymbol{\mu} \in L$ gælder. Det

vurderer ikke på nogen måde om H_1 er i overensstemmelse med data - tværtimod bruger vi jo modeltilpasningen til H_1 som mål for σ^2 , dvs til at vurdere $D(p_1(\mathbf{y}); p_2(\mathbf{y}))$.



Figur 3.1: Test af hypotesen $H_2 : \mu \in L_2$ under antagelse af $H_1 : \mu \in L$

3.6.3 Testet udtrykt i en parametriseret model

Ovenstående test var formuleret koordinatfrit, nemlig udtrykt i underrum og projektioner. Vi vil her supplere fremstillingen ved at formulere testet i

en parametriseret model, dvs udtrykt ved en modelmatrix og de tilsvarende koefficienter.

Antag således, at underrummet L er fastlagt ved modelmatricen $\mathbf{X}_1$ med de m_1 lineært uafhængige søjler $\mathbf{x}_1, \dots, \mathbf{x}_{m_1}$, dvs at modellen H_1 kan udtrykkes som $\boldsymbol{\mu} = \mathbf{X}_1\boldsymbol{\beta}$ for $\boldsymbol{\beta} \in \mathbb{R}^{m_1}$. Estimatet (under H_1) for $\boldsymbol{\beta}$ er derfor (jvf (3.16))

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}_1'\boldsymbol{\Sigma}^{-1}\mathbf{X}_1)^{-1}\mathbf{X}_1'\boldsymbol{\Sigma}^{-1}\mathbf{y} \quad (3.31)$$

og estimatet for σ^2 under H_1 er

$$s_1^2 = \frac{1}{k - m_1} D(\mathbf{y}; \mathbf{X}_1\hat{\boldsymbol{\beta}})$$

Antag nu, at hypotesen H_2 udsiger, at de sidste $m_1 - m_2$ koefficienter, β_i er nul, dvs

$$H_2 : \beta_{m_2+1} = \dots = \beta_{m_1} = 0$$

svarende til $L_2 = \text{span}\{\mathbf{x}_1, \dots, \mathbf{x}_{m_2}\}$.

Betragt nu en opdeling af $\mathbf{X}_1 = \{\mathbf{X}_2, \mathbf{U}\}$, hvor $\mathbf{X}_2$ har søjlerne $\mathbf{x}_1, \dots, \mathbf{x}_{m_2}$ og $\mathbf{U}$ har søjlerne $\mathbf{x}_{m_2+1}, \dots, \mathbf{x}_{m_1}$, og lad tilsvarende $\boldsymbol{\beta}' = \{\boldsymbol{\alpha}', \boldsymbol{\gamma}'\}$, hvor $\boldsymbol{\alpha}$ angiver de første m_2 komponenter af $\boldsymbol{\beta}$, og $\boldsymbol{\gamma}$ angiver de sidste $m_1 - m_2$ komponenter.

I denne parametrisering kan hypotesen altså udtrykkes som

$$H_2 : \boldsymbol{\gamma} = \mathbf{0}$$

svarende til $\boldsymbol{\mu} = \mathbf{X}_2\boldsymbol{\alpha}$ med estimatet for $\boldsymbol{\alpha}$ under H_2

$$\hat{\boldsymbol{\alpha}} = (\mathbf{X}_2'\boldsymbol{\Sigma}^{-1}\mathbf{X}_2)^{-1}\mathbf{X}_2'\boldsymbol{\Sigma}^{-1}\mathbf{y}$$

med den tilsvarende fittede værdi, $\hat{\boldsymbol{\mu}}$ under H_2

$$\hat{\boldsymbol{\mu}} = \mathbf{X}_2\hat{\boldsymbol{\alpha}}$$

Tilsvarende kan estimatet (3.31) betegnes $\hat{\boldsymbol{\beta}}' = \{\hat{\boldsymbol{\alpha}}', \hat{\boldsymbol{\gamma}}'\}$, sådan at den fittede værdi under modellen H_1 kan skrives

$$\hat{\boldsymbol{\mu}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}_2\hat{\boldsymbol{\alpha}} + \mathbf{U}\hat{\boldsymbol{\gamma}}$$

Teststørrelsen (3.28) for H_2 under H_1 er da

$$F(\mathbf{y}) = \frac{D(\hat{\boldsymbol{\mu}}; \hat{\boldsymbol{\mu}})/(m_1 - m_2)}{s_1^2} \quad (3.32)$$

hvor

$$D(\hat{\boldsymbol{\mu}}; \hat{\boldsymbol{\mu}}) = D(\mathbf{X}_2\hat{\boldsymbol{\alpha}} + \mathbf{U}\hat{\boldsymbol{\gamma}}; \mathbf{X}_2\hat{\boldsymbol{\alpha}}) = \|\mathbf{X}_2(\hat{\boldsymbol{\alpha}} - \hat{\hat{\boldsymbol{\alpha}}}) + \mathbf{U}\hat{\boldsymbol{\gamma}}\|$$

jvf (3.6).

I almindelighed vil teststørrelsen for $H_2 : \boldsymbol{\gamma} = \mathbf{0}$ altså afhænge hvilke øvrige parametre $\boldsymbol{\alpha}$, der er i modellen.

Bemærkning 1 *Specialtilfældet med ortogonal parametrisering*

Hvis nu specielt $\mathbf{X}_2'\mathbf{U} = \mathbf{0}$, (dvs hvis rummet udspændt af søjlerne i $\mathbf{U}$ er ortogonalt på L_2), gælder, at $\mathbf{X}_2\hat{\boldsymbol{\alpha}}$ er projektionen ned på L_2 , dvs at $\hat{\boldsymbol{\alpha}} = \hat{\hat{\boldsymbol{\alpha}}}$, uafhængigt af $\hat{\boldsymbol{\gamma}}$. I dette tilfælde får man

$$D(\hat{\boldsymbol{\mu}}; \hat{\boldsymbol{\mu}}) = \|\mathbf{U}\hat{\boldsymbol{\gamma}}\|$$

og F-teststørrelsen (3.28) simplificeres til et test alene baseret på $\hat{\boldsymbol{\gamma}}$

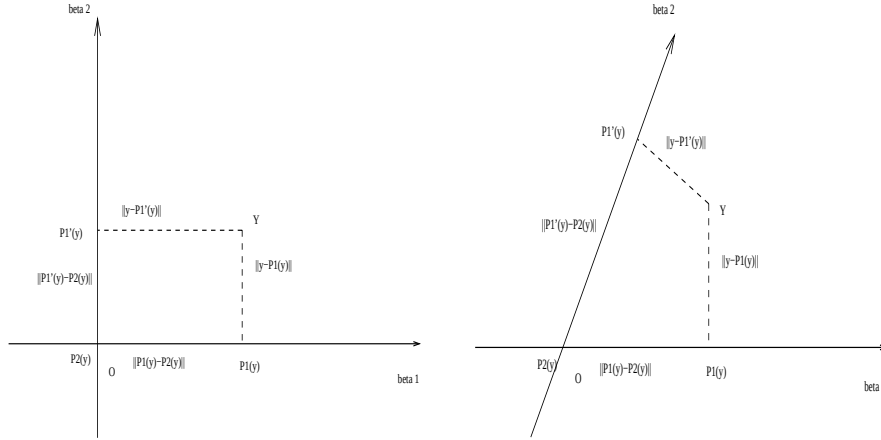
$$F(\mathbf{y}) = \frac{\|\mathbf{U}\hat{\boldsymbol{\gamma}}\|/(m_1 - m_2)}{s_1^2} = \frac{(\mathbf{U}\hat{\boldsymbol{\gamma}})'\boldsymbol{\Sigma}^{-1}\mathbf{U}\hat{\boldsymbol{\gamma}}/(m_1 - m_2)}{s_1^2}$$

Såfremt rummet udspændt af søjlerne svarende til de parametre, der testes, er ortogonalt på rummet L_2 svarende til de øvrige parametre i modellen, vil teststørrelsen for modelreduktion ikke afhænge af hvilke parametre, der i øvrigt er i modellen.

▽

3.6.4 Modelreduktion ved partielle tests

Vi har altså et princip for test for reduktion af en model, nemlig test af en hypotese, der er formuleret som et underrum i en given model. Ofte vil der være flere måder, hvorpå en given model H_i kan udtrykkes som en følge af underrum, der omslutter hinanden. På et givet trin kan man således fortsætte reduktionen ad flere forskellige grene. For at vælge, ad hvilken gren, man vil fortsætte, benyttes sædvanligvis et partielt test:



Figur 3.2: Projektioner ved hhv ortogonale og ikke-ortogonale søjler i $\mathbf{X}$ -matricen

Definition 3.6.1 *Partielt kvotienttest*

Antag at hypotesen H_i har de forskellige delhypoteser $H_{i+1}^A \subset H_i$; $H_{i+1}^B \subset H_i$; $\dots H_{i+1}^K \subset H_i$.

Ved det partielle kvotienttest for H_{i+1}^J under H_i forstås det betingede test for hypotesen H_{i+1}^J under antagelse af at hypotesen H_i kan opretholdes.

Det følger af sætning 3.6.1, at tælleren i F-teststørrelsen i dette partielle test bestemmes som afvigelsen mellem estimerne $\hat{\mu}$ under H_i og estimerne $\hat{\hat{\mu}}$ under H_{i+1}^J

$$F(\mathbf{y}) = \frac{D(\hat{\mu}; \hat{\hat{\mu}})/(m_1 - m_2)}{s_1^2}, \quad (3.33)$$

▽

hvor s_1^2 angiver residualvariansen under H_i .

Bemærkning 1 *Det partielle test betegnes undertiden et type III test*

Det partielle test af H_{i+1}^J under H_i kaldes undertiden det marginale kvotienttest af H_{i+1}^J (under H_i)

I SAS procedurerne kaldes testet ofte et Type III test.

▽

Bemærkning 2 *Det partielle test “korrigerer” for alle de effekter, der er i modellen på det pågældende trin*

Ved tolkning af teststørrelserne svarende til et givet trin i modelhierarkiet siger man, at man vurderer effekten af det givne led korrigeret for de øvrige effekter, der er i modellen på det pågældende trin. ∇

3.6.5 Successiv testning, type I opspaltning

Nu er princippet klart.

Betragt en hierarkisk organiseret kæde af lineære hypoteser, dvs. hypotese-kæder af formen:

$$H_M \subset \dots \subset H_2 \subset H_1 \subset H_0 \subset H_F \quad (3.34)$$

dvs.

$$H_M \Rightarrow \dots \Rightarrow H_2 \Rightarrow H_1 \Rightarrow H_0 \Rightarrow H_F$$

hvor H_i er en lineær hypotese, og hvor H_F angiver den fulde model. Bunden af kæden, H_M , angiver den minimale model, der overhovedet kan komme i betragtning (Sædvanligvis *total homogenitet*, dvs. at alle middelværdier er ens).

En hypotese H_i specificerer et underrum $L_i \subset \mathbb{R}^k$.

Modelhierarkiet svarende til hypoteserne kan da udtrykkes ved følgen af lineære underrum

$$\mathbb{R} \subseteq L_M \dots \subset L_2 \subset L_1 \subset L_0 \subset \mathbb{R}^k, \quad (3.35)$$

Vi har tidligere set, at sådanne hypoteser kan repræsenteres ved en modelmatrix.

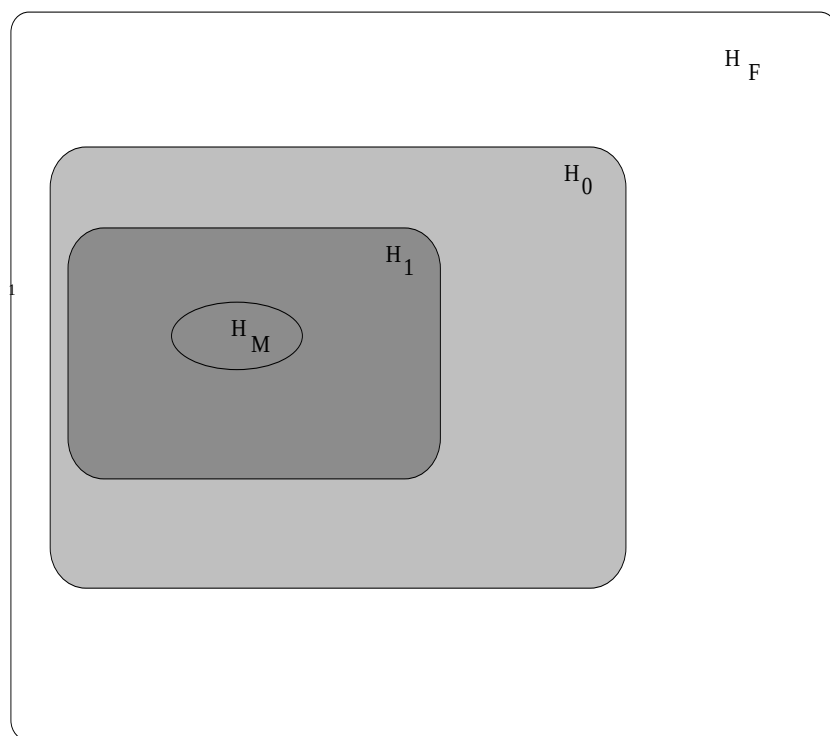
Modelmatricen svarende til den fulde model H_F er den k -dimensionale enhedsmatrix $\mathbf{I}_k$. Lad $\mathbf{X}_i$ angive en modelmatrix svarende til H_i . Den minimale model, H_M , vil sædvanligvis være en model svarende til fuldstændig homogenitet, dvs en model med blot et interceptled. Modelmatricen for denne minimale model er en søjle med lutter ettaller.

Modelhierarkiet svarende til hypoteserne kan tilsvarende udtrykkes ved rummene udspændt af modelmatricerne

$$\mathbb{R} \subset \dots \subset \text{span}(\mathbf{X}_2) \subset \text{span}(\mathbf{X}_1) \subset \text{span}(\mathbf{X}_0) \subset \mathbb{R}^k, \quad (3.36)$$

hvor $\text{span}(\mathbf{X})$ angiver underrummet udspændt af søjlerne i matricen $\mathbf{X}$.

Modelhierarkiet er illustreret i nedenstående figur



Det følger ved gentagen anvendelse af sætning 3.6.1, at

såfremt en kæde af hypoteser er organiseret hierarkisk, kan man spalte residualafvigelsen $D(\mathbf{y}; p_M(\mathbf{y}))$ omkring den minimale model i en sum af bidrag, der hver for sig kan tillægges en selvstændig fortolkning som ændringen i residualkvadratsummen når man reducerer modellen til den nærmest lavere model.

tælleren i F-teststørrelsen svarende til kvotientteststørrelsen for et betinget test.

Den sædvanlige variansanalyseopspaltning, Type I test, udtrykker successive projektioner ned igennem en kæde af underrum svarende til rækkefølgen i modelformlen, hvor der på ethvert trin er korrigeret for de effekter, der står tidligere i modelformlen (højere oppe i variansanalyse-skemaet).

Derimod udtrykker leddene i type III analysen bidraget fra det pågældende led, når **alle de øvrige led** bibeholdes i modellen.

Kun hvis $\mathbf{X}$ matricen er ortogonal svarer type I og type III til hinanden.

I praksis, i situationer, hvor man ikke har en ortogonal opspaltning af modelmatricen (parameterrummet) vil man derfor ofte starte med en stor model - tænke over det tilsvarende inklusionsdiagram (dvs forskellige hypotesekæder af formen (3.34) - og bruge et partielt test (type III) til - under skyldig hensyntagen til analyseformålet - at afgøre hvilken sti, man vil benytte (dvs hvilket led, man vil fjerne). Derefter reestimeres modellen - for at opnå et nyt estimat for residualvariansen, σ^2 (som har flere frihedsgrader, og som afspejler den reelle variation omkring den valgte model), og der foretages modelkontrol for at vurdere om modellen er rimelig, og så fortsætter man på samme måde.

Evt opdager man, at det havde været bedre at vælge en anden sti på et foregående trin, og så går man tilbage i hierarkiet til dette trin, og fortsætter nedad den nye sti.

3.7 Repræsentation af modeller, modelformler og analytiske udtryk

Hidtil har vi i store og hele repræsenteret modeller og delmodeller ved *affine underrum* af $\mathbb{R}^k$, og udtrykt estimation og test ved projektioner og afstande relateret til disse underrum. Dette har været fordelagtigt med henblik på en beskrivelse af de generelle egenskaber, idet disse egenskaber ikke afhænger af de vilkårlige parametriseringer (koordinatsystemer) af rum og underrum. I det omfang, det har

været relevant, har vi specificeret et underrum ved en *modelmatrix* uden dog at gå nærmere ind på formuleringen af modelmatricen.

I praksis vil man som regel benytte en eller anden form for parametrisering af problemet for lettere at kunne kommunikere resultatet af analysen. Vi vil her beskrive forskellige repræsentationer af generelle lineære modeller, nemlig

repræsentation ved modelmatrix: En repræsentation

$$\mu = \mathbf{X}\beta$$

ved et sæt af vektorer, der udspænder det lineære underrum

Parametrisk repræsentation: En repræsentation ved et regneudtryk, fx

$$\mu_i = \beta_0 + \beta_1 x_{ij}$$

(sædvanligvis byggende på koordinater i et lineært underrum). Denne repræsentation benyttes sædvanligvis i lærebøger og i rapportering af statistiske analyser.

Repræsentation ved modelformel: En repræsentation ved symboler for de *variable*, der indgår i modellen, fx

$$\text{Ramus} = \text{Alder} .$$

Modelformlerne benyttes til formulering af lineære modeller i statistikprogrammer som fx SAS, R og S-plus

I sædvanligt statistikprogrammer formuleres modeller ved en *modelformel*, der beskriver responsstrukturen udtrykt ved de *forklarende variable*, også kaldet *kovariabel*. Output refererer til koefficienter i en fortolkning af denne modelformel svarende til en (undertiden singulær) modelmatrix.

I det følgende vil vi beskrive, hvorledes modelmatricen svarende til henholdsvis en *kontinuert kovariabel* og en *kvalitativ kovariabel* fremkommer, og vi vil beskrive de tilsvarende parametriske repræsentationer. Endvidere vil vi beskrive, hvordan de variable sammensættes i mere komplicerede modelformeludtryk.

Parametrene i den parametriske repræsentation af en hypotese kan opfattes som “koordinater” i det lineære underrum, L , der svarer til den pågældende hypotese, i et koordinatsystem med vektorerne i modelmatricen som basisvektorer..

Enhver repræsentation af de forklarende variable ved en $k \times m$ matrix

$$\mathbf{X} = (\mathbf{x}_1 | \mathbf{x}_2 | \cdots | \mathbf{x}_m)$$

modsvares af en parametrisering af underrummet $L = \text{span}(\mathbf{X})$ sådan at ethvert $\mu \in L$ kan udtrykkes som en linearkombination

$$\mu_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_m x_{im}, \quad i = 1, 2, \dots, k.$$

Matricen $\mathbf{X}$ kaldes *modelmatricen svarende til de m kovariable*.

3.7.1 Intercept led

Sædvanligvis vil man i modelleringen inddrage et såkaldt intercept led, der specificerer den *minimale model*, nemlig modellen svarende til en fælles middelværdi for alle observationer. I det sædvanlige variansanalyseskema er det jo netop variationen omkring det fælles gennemsnit (Corrected Sum og Squares), der opspaltes i bidrag fra hhv **Model** og **Error**.

I modelmatricen indgår interceptleddet som den første søjle, bestående af lutter ettaller,

$$\mathbf{X} = \mathbf{1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \quad (3.37)$$

I en parametrisk fremstilling

$$\mu_i = \beta_0 + \sum_{j=1}^m \beta_j x_{ij}, \quad i = 1, 2, \dots, k.$$

er interceptleddet størrelsen β_0 , og den minimale model fremkommer netop for

$$\beta_1 = \beta_2 = \cdots = \beta_k = 0$$

I modelformler i R og S-plus indgår interceptleddet sædvanligvis implicit. Såfremt man ønsker eksplicit at referere til interceptleddet, beskrives det i modelformlen ved et ettal.

I modelformler i SAS indgår interceptleddet implicit. Såfremt man ønsker en model uden interceptled vælges i SAS-procedurer en option **NOINT** efter sætningen, der angiver modellen. I SAS Insight markeres **No Intercept** i Model-vinduet.

3.7.2 Kontinuerte kovariabel

En kontinuert kovariabel er en kovariabel, hvis værdier udtrykkes på en måleskala med reelle værdier, en såkaldt *intervalskala*, som længde, vægt, hastighed, koncentration, antal etc.

Betragt et observationssæt med de observerede værdier $y_1, y_2, \dots, y_k$ af interessevariablen og med de tilhørende værdier $x_1, x_2, \dots, x_k$ af en kontinuert kovariabel.

Det parametriske udtryk

$$\mu_i = \beta_0 + x_i \beta_1,$$

hvor $\beta_0 \in \mathbb{R}$ og $\beta_1 \in \mathbb{R}$ angiver den sædvanlige regressionsanalysemodel med én forklarende variabel.

Den tilsvarende modelmatrix er

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \\ 1 & x_k \end{pmatrix},$$

og såfremt responserne y_i er placeret i en variabel med navnet $\mathbf{Y}$ og værdierne x_i af den forklarende variabel er placeret i en variabel med navnet $\mathbf{X}$, er den tilsvarende modelformel

$$\mathbf{Y} = \mathbf{X}$$

Sammensætning af flere kontinuerte kovariabel

For en model med m kontinuerte kovariabel vil værdierne af hver af de m kovariabel kunne repræsenteres ved en k -dimensional søjlevektor, $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$, hvor $\mathbf{x}_j$ angiver værdierne af den j 'te kovariabel

$$\mathbf{x}_j = \begin{pmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{kj} \end{pmatrix} \quad (3.38)$$

Den tilsvarende modelmatrix er

$$\mathbf{X} = (\mathbf{1} | \mathbf{x}_1 | \mathbf{x}_2 | \dots | \mathbf{x}_m)$$

hvor $\mathbf{1}$ angiver en søjle af ettaller jvf (3.37). Modelmatricen svarer til en parametrisering af underrummet $L = \text{span}(\mathbf{X})$ sådan at ethvert $\mu \in L$ kan udtrykkes ed

det parametriske udtryk

$$\mu_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_m x_{im}, \quad i = 1, 2, \dots, k. \quad (3.39)$$

Såfremt responserne y_i er placeret i en variabel med navnet Y og værdierne af de forklarende variable er placeret i variable med navnene $X_1, X_2, \dots, X_m$, vil den tilsvarende modelformel i SAS være

$$Y = X_1 \quad X_2 \quad \dots \quad X_m$$

og i R og S-Plus er formelen

$$Y = X_1 + X_2 + \dots + X_m$$

Prikkerne ... skrives ikke, men de symboliserer led af formen XP .

En modelformel i SAS af formen

$$Y = X_1 \quad X_2 \quad X_1 * X_2$$

svarer i R og S-plus til formelen

$$Y = X_1 + X_2 + X_1 * X_2$$

Den tilsvarende parametriske repræsentation er

$$\mu_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} x_{i2}$$

Specielt vil modelformlen

$$Y = X_1 * X_1$$

således modsvare den parametriske model

$$\mu_i = \beta_0 + \beta_1 x_{i1}^2$$

dvs en lineær model i x_1^2 .

3.7.3 Kvalitative kovariable, faktorvariable

Kvalitative kovariable er variable, hvis værdier udtrykkes på en *nominal skala*. Kvalitative variable udtrykker resultatet af en *klassifikation*. Sådanne kovariable betegnes ofte *faktorvariable*. En faktor svarer til en klassifikation af indeksmængden i disjunkte grupper. En faktor kan kun have et begrænset antal værdier, de

så kaldte *niveauer* (eng. *levels*). For en faktor med r niveauer kan vi altid repræsentere niveauerne ved en *liste* af heltal $1, 2, \dots, r$, de *formelle niveauer*. I praksis vil der ofte være knyttet navne, eller evt en numerisk kodning til de formelle niveauer.

For en faktor med r niveauer kan man for hvert niveau konstruere en k -dimensional søjlevektor $\mathbf{u}_j$, den så kaldte *incidensvektor*. Vektoren består af nuller og ettaller og har et ettal på den i 'te plads, hvis den i 'te observation er foretaget ved faktorniveau j , ellers et nul.

Incidensmatricen $\mathbf{U} = \{\mathbf{u}_1 | \mathbf{u}_2 | \dots | \mathbf{u}_r\}$ udtrykker således allokeringen af de r faktorniveauer til de enkelte observationer. Rækkesummen i $\mathbf{U}$ er 1, da hver observation er tilknyttet ét og kun ét faktorniveau.

En model med én kvalitativ kovariabel kan udtrykkes ved modelmatricen $\mathbf{U}$ som

$$\boldsymbol{\mu} = \mathbf{U}\boldsymbol{\alpha}$$

svarende til den parametriske fremstilling

$$\mu_i = \sum_j u_{ij} \alpha_j = \alpha_{a(i)} , \quad i = 1, 2, \dots, k , \quad (3.40)$$

hvor $a(i)$ angiver det formelle faktorniveau svarende til den i 'te observation.¹

Inddrages et interceptled (svarende til den minimale model) i modellen, fås en repræsentation ved en modelmatrix

$$\mathbf{X} = \{\mathbf{1}, \mathbf{u}_1, \dots, \mathbf{u}_r\}$$

svarende til en parametrisk fremstilling

$$\mu_i = \beta_0 + \sum_j u_{ij} \alpha_j = \beta_0 + \alpha_{a(i)} , \quad i = 1, 2, \dots, k , \quad (3.41)$$

hvor parametrene $\alpha_1, \dots, \alpha_p$ i fremstillingen (3.41) må underkastes nogle bånd for at tilgodese overparametriseringen.

¹Hvis der er flere gentagelser for hvert faktorniveau kan man opstille data på tabelform. Således kan man - i situationen med en enkelt faktor - opstille data i et skema med r rækker (svarende til de r faktorniveauer) og på hver af de r linier anføre de observerede værdier svarende til dette faktorniveau. Den i 'te observation bliver da placeret i p 'te række, hvor $p = a(i)$ angiver det formelle niveau af faktoren for den i 'te observation. Observationerne indiceres da som $y_{p,j}$, hvor $j = 1, 2, \dots, n_p$ angiver gentagelsesnummeret ved den pågældende faktorværdi.

Når data repræsenteres på denne form er den tilsvarende parametriske model på formen $\mu_p = \alpha_p$, $p = 1, 2, \dots, r$ eller, hvis man inddrager et interceptled: $\mu_p = \beta_0 + \alpha_p$, $p = 1, 2, \dots, r$

Parametrisering ved kontraster

De bånd, der ofte indføres, er såkaldte *kontraster*, nemlig størrelser, der direkte kan estimeres.

Problemet er, at matricen

$$\mathbf{X}_A = \{\mathbf{1}, \mathbf{u}_1, \dots, \mathbf{u}_r\}$$

svarende til repræsentationen (3.41) kun har rangen r . Én af vektorerne $\mathbf{u}_1, \dots, \mathbf{u}_r$ kan udtrykkes som en linearkombination af de øvrige og vektoren $\mathbf{1}$. Vi ved jo netop, at rækkesummen af $\mathbf{U}_A$ er 1.

Man må derfor indføre et lineært bånd mellem parametrene $\beta_1, \dots, \beta_r$, eller *reparametrisere* problemet.

Såfremt man ønsker at operere med en modelmatrix af fuld rang, kan man vælge at reparametrisere problemet sådan at man opnår en $k \times r$ dimensional matrix

$$\mathbf{X}_A = \{\mathbf{1}, \mathbf{c}_1, \dots, \mathbf{c}_{r-1}\}$$

af fuld rang, der udspænder L_A .

Vektorerne $\mathbf{c}_j$ definerer en række såkaldte *estimable kontraster* imellem de r størrelser $\beta_1, \dots, \beta_r$

Der findes en række forskellige muligheder for at definere kontraster mellem r niveauer af en faktor. I nedenstående eksempel skal vi illustrere nogle af disse muligheder i tilfældet $r = 4$.

Eksempel 3.7.1 Kontraster mellem $r = 4$ niveauer

Helmert-transformation

Parametrisering ved den såkaldte Helmert-transformation svarer til modellen

$$\boldsymbol{\mu} = \mathbf{X}_H \begin{pmatrix} \beta_0 \\ \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{pmatrix}$$

hvor modelmatricen $\mathbf{X}_H$ er

$$\mathbf{X}_H = \begin{pmatrix} 1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & 0 & 2 & -1 \\ 1 & 0 & 0 & 3 \end{pmatrix},$$

Parametriseringen svarer til at første kontrast angiver forskellen $\beta_2 - \beta_1$ mellem værdien svarende til niveau 2 og niveau 1; anden kontrast angiver forskellen $\beta_3 -$

$(\beta_1 + \beta_2)/2$ mellem værdien svarende til niveau 3 og gennemsnittet mellem niveau 1 og niveau 2. Den j 'te kontrast angiver forskellen $\beta_{j+1} - (\beta_1 + \dots + \beta_j)/j$ mellem niveau $j + 1$ og gennemsnittet af niveauerne 1 til j .

Helmert-transformationen sikrer, at de enkelte søjler $\mathbf{c}_j$ er indbyrdes ortogonale, og endvidere ortogonale på vektoren $\mathbf{1}$.

Sum-kodning

En anden mulighed er den såkaldte “sum”-kodning

$$\boldsymbol{\mu} = \mathbf{X}_S \begin{pmatrix} \beta_0 \\ \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{pmatrix}$$

hvor modelmatricen $\mathbf{X}_S$ er

$$\mathbf{X}_H = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & -1 & -1 & -1 \end{pmatrix},$$

svarende til at første kontrast angiver forskellen mellem β_1 og β_r , anden kontrast angiver forskellen mellem β_2 og β_r , etc. Også disse kontraster er indbyrdes ortogonale og ortogonale på $\mathbf{1}$.

Treatment-kodning

Ofte benytter man en kodning svarende til repræsentationen

$$\boldsymbol{\mu} = \mathbf{X}_T \begin{pmatrix} \beta_0 \\ \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{pmatrix}$$

hvor modelmatricen $\mathbf{X}_T$ er

$$\mathbf{X}_T = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix},$$

Kodningen kaldes undertiden “treatment”-kodning. SAS benytter sævanligvis en variant af denne kodning, nemlig med det sidste niveau som referenceniveau.

Reparametriseringen svarer til at benytte det første niveau β_0 som referenceniveau. Parametrene $\gamma_1, \dots, \gamma_{r-1}$ svarende til denne kodning er ikke kontraster, da søjlesummerne ikke er nul, og søjlerne derfor ikke er ortogonale på vektoren $\mathbf{1}$. ∇

3.7.4 Modelformler

I mange programsystemer benyttes en kompakt *modelformel* til specifikation af lineære modeller.

Vi vil her skitsere en symbolsk notation, der er foreslået af Wilkinson og Rogers [18]. Notationen bruges i det store og hele i sædvanligt statistikprogrammel, f.eks. R og S-plus, hvor man refererer til variable ved symbolske variabelnavne. SAS bruger en variant af notationen (fx. benyttes prik-operatoren ikke, og plus operatoren er erstattet af et mellemrum).

En *modelformel* er en symbolsk beskrivelse af middelværdistrukturen i en lineær model ved en kombination af symboler for de variable og *operatorer*.

Vi vil nedenfor beskrive disse operatorer ved hjælp af de tilknyttede operationer på de tilsvarende modelmatricer og de led, der genereres i den parametriske repræsentation.

Vi vil lade A og B betegne faktorvariable (klassifikationer af indeksmængden), og X angive en kontinuert kovariabel. Lad incidensmatricerne for A og B være hhv $\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r)$ for A , og $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_s)$ for B , og lad X have modelvektoren $\mathbf{x}$

Prik- eller punktopoperatoren

Operatoren skrives ofte som et punktum “.” eller kolon “:”. Operatoren genererer led i den parametriske repræsentation svarende til alle kombinationer af de to variable, der “prikkes”.

For to faktorvariable A og B dannes den $n \times (r \cdot s)$ -dimensionale incidensmatrix svarende til $A.B$ ved elementvis multiplikation af alle søjlepar $\mathbf{u}_p$ og $\mathbf{w}_q$. Leddene i den parametriske repræsentation bliver de $r \cdot s$ led af formen $\gamma_{p,q}$, $p = 1, 2, \dots, r$; $q = 1, 2, \dots, s$.

For en faktorvariabel A og en kontinuert variabel X dannes den $n \times r$ -dimensionale modelmatrix svarende til $A.X$ ved elementvis multiplikation af $\mathbf{x}$ med hver af søjlerne $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r$. Leddene i den parametriske repræsentation på faktorform bliver de r led af formen $\alpha_p x$, $p = 1, 2, \dots, r$.

Der gælder

$$A.A = A$$

svarende til at incidensmatricen for $A.A$ har r søjler lig med søjlerne for A , og de resterende $r(r - 1)$ søjler er nulvektorer.

Almindeligvis vil $X.X \neq X$. Venstre side kan fortolkes som en vektor med komponenter x_i^2 . De fleste programsystemer tillader dog ikke denne konstruktion. Hvis man ønsker en forklarende variabel svarende til X^2 bør denne konstrueres eksplicit.

Prikoperatoren er kommutativ. Der gælder

$$A.B \equiv B.A$$

Den er desuden associativ, idet

$$(A.B).C \equiv A.(B.C)$$

Man kan derfor blot skrive $A.B.C$ etc.

Plusoperatoren

Plusoperatoren i modelformler udtrykker addition af parametre i den parametriske repræsentation. Matricen svarende til plusoperationen fremkommer ved at sætte addendernes matricer ved siden af hinanden, dvs. matricen svarende til $A + B$ fås som $(\mathbf{U}|\mathbf{W})$ og matricen svarende til $A + X$ fås som $(\mathbf{U}|\mathbf{x})$.

Det rum, der udspændes af incidensmatricen svarende til $A + B$ er således summen af rummene udspændt af modelmatricen svarende til A og incidensmatricen svarende til B .

Der gælder

$$A + A \equiv A$$

For at slippe for at skrive for mange parenteser bruger man almindeligvis den samme konvention som ved de sædvanlige plus og gange operationer, at “+” har lavere prioritet end “.”, dvs.

$$A.B + C \equiv (A.B) + C$$

Prikoperatoren “.” er distributiv med hensyn til plusoperatoren “+”, dvs

$$A.(B + C) \equiv A.B + A.C$$

Krydsede faktorer

Krydsningsoperatoren “*” bruges hovedsagelig til at give en kompakt beskrivelse af modeller med flere faktorvariable. Operatoren “*” defineres som

$$A * B = A + B + A.B$$

$$A * B * C = A + B + C + A.B + A.C + B.C + A.B.C$$

etc. I den parametriske repræsentation svarer modelformlen $A * B$ således til led af formen $\alpha_p + \beta_q + \gamma_{p,q}$, dvs. modellering af hovedvirkningerne af A og B samt af vekselvirkningen mellem A og B .

Mange programsystemer tillader yderligere, at de variable A , B etc. i ovenstående udtryk erstattes med modelformler.

Krydsningsoperatoren “ $*$ ” har højere prioritet end “ $+$ ”, men lavere prioritet end “ $.$ ”, dvs

$$\begin{aligned} A * B + C &\equiv A + B + C + A.B \\ A * B.C &\equiv A + B.C + A.B.C \end{aligned}$$

Krydsningsoperatoren “ $*$ ” er associativ og distributiv med hensyn til “ $+$ ”. Der gælder nemlig

$$\begin{aligned} A * (B + C) &\equiv A + (B + C) + A.(B + C) \\ &\equiv A + B + C + A.B + A.C \\ &\equiv A + B + A.B + A + C + A.C \\ &\equiv A * B + A * C \end{aligned}$$

Fjernelse af led i modelformel

Operatoren “ $-$ ” defineres som den modsatte operation af “ $+$ ”. Operatoren bruges til at udtrykke fjernelse af led i en modelformel.

Man har således udtrykket for en tofaktormodel uden vekselvirkning, dvs. alene med hovedvirkninger

$$A * B - A.B \equiv A + B$$

og tilsvarende udtrykker

$$A * B * C - A.B.C \equiv A + B + C + A.B + A.C + B.C$$

en model med tre faktorer, der indeholder alle tre hovedvirkninger og alle tre to-faktorvekselvirkninger.

Hierarkisk organiseret indeksmængde, underordnede faktorer

Vi nævner kort den særlige struktur, der er knyttet til en hierarkisk klassifikation af indeksmængden. En *hierarkisk klassifikation* optræder når niveauerne for én klassifikation, f.eks. B , er underordnet niveauerne for en anden klassifikation, f.eks. A .

I modelformlen udtrykker man i R og S-plus underordning ved symbolet “ $/$ ”. Således udtrykker

$$A/B = A + A.B$$

at B er underordnet A . Fortolkningen af $A.B$ er her effekter af B indenfor A .² Vi bemærker, at A/B adskiller sig fra $A * B$ ved at der ikke er noget led, der udtrykker hovedvirkning af B . Et sådant led er jo uden mening, da niveauerne af faktor B jo relaterer til det pågældende niveau af faktor A .

Såfremt man tillader symbolerne i modelformlen selv at symbolisere modelformler, defineres “/” ved

$$A/B = A + \text{pl}(A).B$$

hvor vi fortolker $\text{pl}(A)$ som “produktleddet” (ved brug af prikoperatoren) af alle elementerne i A .

Således har vi eksempelvis

$$(A * B)/C \equiv A * B + A.B.C$$

Underordningsoperatoren er associativ. Der gælder således

$$A/(B/C) \equiv (A/B)/C$$

Underordningsoperatoren er distributiv med hensyn til “+” idet der gælder

$$A/(B + C) = A + A.(B + C) = A + A.B + A + A.C = A/B + A/C$$

I lighed med krydsningsoperatoren har underordningsoperatoren en prioritet imellem “.” og “+”. Man vælger ofte konventionen at give den prioritet over “*.”.

3.7.5 Inklusionsdiagrammer for hypotesekæder, Type I og Type III opspaltning

Ved reduktion af en given model vil det sædvanligvis være fordelagtigt at bevæge sig ned igennem en hypotesekæde, hvor de tilsvarende lineære rum ligger inden i hinanden. Vi vil her illustrere sådanne hypotesekæder for nogle simple modelsituationer ved brug af inklusionsdiagrammer.

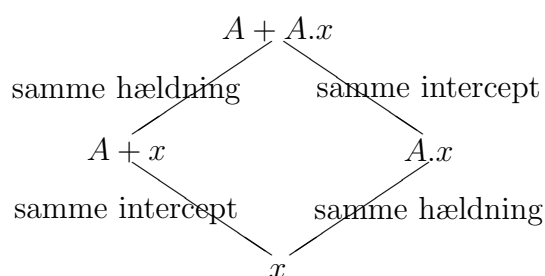
²I programsystemet SAS bruges en parentes til at angive underordning af faktorer. I SAS symboliserer leddet $B(A)$, at faktoren B er underordnet A . Angivelse af leddet $B(A)$ i en modelformel indebærer sædvanligvis, at alle vekselvirkningsleddene mellem B og A indgår i modellen. Såfremt man ønsker hovedeffekterne af A modelleret (af hensyn til deviansopspaltningen), skal man eksplicit angive det i modelformlen ved at skrive $A + B(A)$.

Sammenligning af to (eller flere) regressionslinier

I figur 3.3 vises en situation med to (eller flere) regressionslinier, der kan reduceres til en enkelt regressionslinie ved først at teste for fælles hældning, og derefter for fælles intercept, eller først at teste for fælles intercept og derefter for fælles hældning.

Bemærk at leddet $sfA * x$ i modelformlen svarer til at modellere separate hældninger for hvert af niveauerne i faktoren A

Figur 3.3: Inklusionsdiagram svarende til sammenligning af regressionslinier

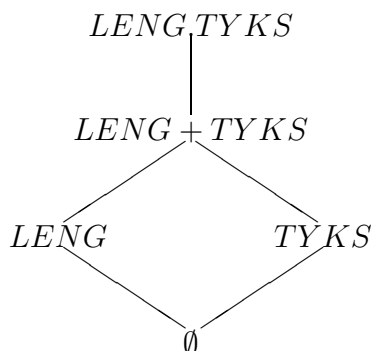


Tofaktorforsøg

I et forsøg med udplantning af blommestiklinger betragtede man to *længder*, henholdsvis 6 [cm] og 12 [cm] og tre *tykkelsesgrader*, 3-6 [mm], 6-9 [mm] og 9-12 [mm]. I oktober udplantede man 20 stiklinger af hver kombination af længde og tykkelse, og året efter vurderede man egenskaberne for de udplantede stiklinger.

Nedenstående inklusionsdiagram figur 3.4 illustrerer de to kæder, der kan konstrueres for et sådant forsøg med to faktorer. Nemlig først et test for forsvindende vekselvirkning, hvorefter man enten kan teste for forsvindende søjleeffekt (TYKS) og derefter ved accept teste for forsvindende søjleeffekt (LENG) til den minimale model - eller man kan først teste for rækkeeffekt og derefter søjleeffekt.

Figur 3.4: Inklusionsdiagram svarende til tofaktormodeller for blommeforsøg



Trefaktorforsøg

Endelig illustreres i figur 3.5 de mulige hypotesekæder i et forsøg med tre kvalitative faktorer.

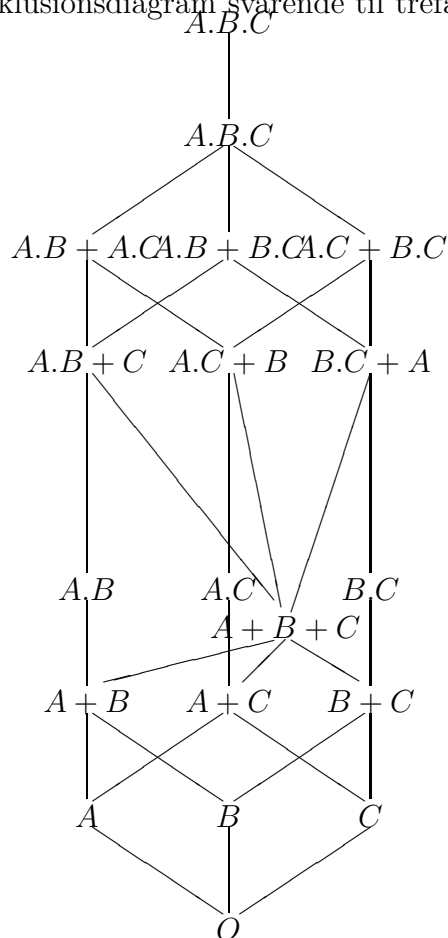
Fremgangsmåde ved modelreduktion

For enhver model, H_0 , kan man formulere mindst én kæde af hierarkisk organiserede hypoteser med den minimale model H_M som den laveste model i hierarkiet, og med H_0 som den højeste model. Principielt kan man så starte med en stor model, H_0 (dog med en dimension mindre end k , sådan at man får residualer forskellige fra nul), og derefter teste (ved brug af sætning 3.6.1) om det er forsvarligt at reducere modellen til den mindre model H_1 . Hvis man vælger at fortsætte med modellen H_1 , kan man da forsøge sig med næste mindre model H_2 . Men det betingede F-test for H_2 under H_1 forudsætter principielt reestimation af σ^2 svarende til residualafvigelsen under H_1 .

Den såkaldte Type I opspaltning af kvadratafgivelsessummer i SAS og S-plus svarer til en opspaltning af $D(\mathbf{y}; p_M(\mathbf{y}))$ i overensstemmelse med denne hypotesekæde. Man skal således huske bemærkningen side 62, at kvadratafgivelsessummen i en type I opspaltning angiver ændringen i D svarende til det pågældende modelled, *når der er korrigeret for den eventuelle effekt af de foranstående led*.

Bidragene i Type I opspaltningen afhænger altså af hvilke foranstående led,

Figur 3.5: Inklusionsdiagram svarende til trefaktormodeller



der er i modellen. Ved tolkning af teststørrelserne svarende til et givet trin i modelhierarkiet siger man, at man vurderer effekten af det givne led *korrigeret for* de øvrige effekter, der er i modellen på det pågældende trin (og F-størrelserne i en Type I opspaltning skal tages med et gran salt, da nævneren jo knytter sig til modellen svarende til alle led).

Som vi så af inklusionsdiagrammet figur 3.3, vil man imidlertid i almindelighed kunne organisere en given model som en sådan kæde på flere forskellige måder. Dog vil vekselvirkningsled være højere end de tilsvarende hovedeffekter.

Det fremgår imidlertid af sætning 4.10.1, at

teststørrelsen for fjernelse af et bestemt led i modelformlen afhænger af, hvilke andre led, der indgår i modellen på det pågældende trin.

I praksis vil man derfor sikre sig at alle relevante led tilgodeses, når man vurderer, hvorvidt et bestemt modelled kan udelades af modellen.

Såfremt man i situationen svarende til figur 3.4 har accepteret hypotesen om forsvindende vekselvirkning, dvs at man har godtaget modellen svarende til modelformlen $LENG + TYK$, da vil testet svarende til den venstre gren i diagrammet modsvare at man undersøger, om der er påviselig effekt af tykkelsen, når man tilgodeser at stiklingerne har forskellig længde og korrigerer herfor. Tilsvarende vil testet svarende til den højre gren i diagrammet modsvare at man undersøger, om der er påviselig effekt af længden, når man tilgodeser at stiklingerne har forskellig tykkelse og korrigerer herfor.

3.8 Eksempel, Blodgennemstrømningsindeks hos malere

Nedenstående data hidrører fra en undersøgelse af effekten af brugen af organiske opløsningsmidler. (Kilde: P.Arlien-Søborg, L.Henriksen, A.Gade, C.Gyldensted: Cerebral blood flow in chronic toxic encephalopathy in house painters exposed to organic solvents. *Acta Neurologica Scandinavica* **66** (1982), pp 34-41).

Hjerneblodgennemstrømningsindex (ISI) tages ofte som udtryk for hjerneaktiviteten. Ved undersøgelsen målt ISI hos 20 malere, hvoraf 9 havde benyttet organiske opløsningsmidler i deres arbejde, og 11 (kontrolgruppen) ikke havde benyttet organiske opløsningsmidler.

```
* kursus/stat3/malblod.sas ;
OPTIONS LINESIZE=80 PAGESIZE = 65;
DATA BLOD;
TITLE 'Blodgennnemstromning hos malere';
INPUT NR ISI ALDER ORG$;
CARDS;
1 48 23 ORG
2 41 35 ORG
```

```

3  36 37 ORG
4  35 38 ORG
5  43 40 ORG
6  44 40 ORG
7  28 45 ORG
8  28 52 ORG
9  31 58 ORG
1  55 31 EJ
2  51 32 EJ
3  44 34 EJ
4  45 38 EJ
5  48 38 EJ
6  50 38 EJ
7  49 46 EJ
8  34 48 EJ
9  31 53 EJ
10 33 57 EJ
11 44 39 EJ
;
PROC SORT ;
BY ORG;
PROC PRINT;
PROC MEANS ;
BY ORG;
VAR ISI ALDER;
RUN;

```

Modelmatricen for en model svarende til forskellig aldersafhængighed og forskelligt intercept for de to grupper er

1	1	0	23	23	0
1	1	0	35	35	0
⋮	⋮	⋮	⋮	⋮	⋮
1	1	0	58	58	0
1	0	1	31	0	31
⋮	⋮	⋮	⋮	⋮	⋮
1	0	1	39	0	39

idet vi har valgt (i lighed med SAS) at overparametrisere modellen således at der er parametrene $\beta_1 = \text{intercept}$, β_2 (opløsningsmiddel), β_3 (ej opløsningsmiddel) β_4

(koefficient til alder), β_5 (koefficient til alder for opløsn.) β_6 (koefficient til alder for ej opløsning)

Øvelse

Tegn et boxplot for ISI for de to grupper

Tastetryk Boxplot → ISI som Y , ORG som X .

Kan man påvise forskel ??

Prøv dernæst model for sammenligning af middelværdi i to stikprøver (t-test for to uafhængige stikprøver).

Tastetryk Fit → ISI som Y , ORG som X .

Kan man påvise forskel ??

Tegn ISI mod alder

Tastetryk Scatterplot → ISI som Y , ALDER som X .

Marker evt de to grupper hver for sig.

Det synes som om ISI aftager med alderen. (Et velkendt fænomen :-))

Og de to grupper har ikke helt samme aldersmæssige sammensætning.

Man bør nok korrigere for alderseffekten ved sammenligningen, fx ved at bruge en model af formen:.

MODEL ISI = ORG ALDER ORG*ALDER

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	960.3088	320.1029	15.18	<.0001
Error	16	337.4912	21.0932		
Corrected Total	19	1297.8000			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
ORG	1	234.9111	234.9111	11.14	0.0042
ALDER	1	705.7162	705.7162	33.46	<.0001
ALDER*ORG	1	19.6815	19.6815	0.93	0.3484

Source	DF	Type III SS	Mean Square	F Value	Pr > F
ORG	1	58.9495	58.9495	2.79	0.1140
ALDER	1	715.9166	715.9166	33.94	<.0001
ALDER*ORG	1	19.6815	19.6815	0.93	0.3484

Nedenfor angives udskrifterne af Type I og Type III svarende til en rækkefølge `alder org` i modelformlen. Bemærk ændringen i type I afvigelser.

Source	DF	Type I SS	Mean Square	F Value	Pr > F
ALDER	1	687.8222	687.8222	32.61	<.0001
ORG	1	252.8050	252.8050	11.99	0.0032
ALDER*ORG	1	19.6815	19.6815	0.93	0.3484

Variable	ORG	DF	Estimate	Std Error	t Stat	> t
Intercept		1	60.3445	6.7622	8.92	<.0001
ORG	EJ	1	16.4294	9.8277	1.67	0.1140
	ORG	0	0.0000	.	.	.
ALDER		1	-0.5682	0.1611	-3.53	0.0028
ALDER*ORG	EJ	1	-0.2259	0.2338	-0.97	0.3484
	ORG	0	0.0000	.	.	.

3.9 Vurdering af modeltilpasning, diagnostiske størrelser

3.9.1 Residualer, standardisering og studentisering

Vi så i det foregående afsnit (3.4), at variansen (dvs usikkerheden) på residuallet blandt andet afhang af det tilsvarende diagonalelement i hat-matricen.

Det vil være naturligt at skalere residualerne med spredningen for at sikre en sammenlignelighed af residualerne. For de generaliserede lineære modeller siger man sædvanligvis at residualerne er *standardiserede*, hvis de er skaleret med en størrelse, der sikrer, at de har samme varians under modellen. Der er imidlertid ikke fuldstændig enighed om denne sprogbrug. Det er jo også en udbredt sprogbrug at bruge betegnelsen “standardiseret” om en variabel, hvis fordeling har middelværdi nul og variansen 1.

For den generelle lineære model indgår den ukendte varians, σ^2 i variansen på residualet, så derfor må man erstatte σ^2 med estimatet. I den generelle lineære model (normalfordeling) i SAS-insight betegner et *standardiseret residual*, residualet divideret med estimatet for dets standardafvigelse.

Imidlertid indgår det pågældende residual jo i beregningen af estimatet for σ^2 , og hvis man vil vurdere størrelsen af et residual i forhold til den estimerede spredning på dette residual, kan det være ønskeligt, at den estimerede spredning er *uafhængig* af netop den betragtede observation. Man siger, at et sæt residualer er *studentiserede*, hvis de er skaleret med et uafhængigt estimat for spredningen.

Nedenstående angives udtryk for de standardiserede og studentiserede residualer, der sædvanligvis optræder i programmer til analyse af generaliserede lineære modeller, som f.eks. programsystemerne SAS og S-Plus.

I udtrykkene betegner h_{ii} det i 'te diagonalelement i hat-matricen, $\hat{\sigma}^2$ angiver estimatet for variansen σ^2 ved benyttelse af alle observationer, og $\hat{\sigma}_{(i)}^2$ angiver estimatet, beregnet ved udeladelse af den i 'te observation.

Standardiseret residual

$$r_i^{rs} = \frac{r_i^r}{\sqrt{\hat{\sigma}^2(1 - h_{ii})}}$$

Studentiseret residual

$$r_i^{rt} = \frac{r_i^r}{\sqrt{\hat{\sigma}_{(i)}^2(1 - h_{ii})}}$$

Under normalfordelingsantagelsen gælder, at det studentiserede residual følger en Student's t-fordeling med $f = k - m - 1$, idet estimatet, $\hat{\sigma}_{(i)}^2$ følger en

$\sigma^2 \chi^2(f)/f$ fordeling og er uafhængig af r_i . (Se fx Jørgensen, linear models).

3.9.2 Kontrol af enkeltobservationer, leverage

Diagonalelementerne i hat-matricen er udtryk for den “vægt” hvormed den enkelte observation bidrager til den fittede værdi for den pågældende datapunkt.

Dette følger direkte ved at betragte

$$\frac{\partial \hat{\boldsymbol{\mu}}}{\partial \mathbf{y}} = \mathbf{H}$$

hvor $\mathbf{H}$ angiver hat-matricen.

Diagonalelementerne h_{ii} i hat-matricen $\mathbf{H}$ angiver således ændringen i den fittede værdi $\hat{\mu}_i$ ved en ændring i y_i .

Eksempel 3.9.1 *Leverage i en enkel regressionsmodel*

For at illustrere begrebet leverage vil vi betragte en regressionsmodel svarende til normalfordelingen, dvs en model med en konstant varians.

Betragt modelmatricen svarende til et intercept-led og én uafhængig variabel

$$\mathbf{X} = \begin{pmatrix} 1 & (x_1 - \bar{x}.) \\ 1 & (x_2 - \bar{x}.) \\ \vdots & \vdots \\ 1 & (x_k - \bar{x}.) \end{pmatrix},$$

hvor vi har ortogonaliseret de to søjler ved at subtrahere gennemsnitsværdien af x_i fra samtlige x -værdier. Vi har da

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} k & 0 \\ 0 & \sum (x_i - \bar{x}.)^2 \end{pmatrix}$$

hvorfor

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} 1/k & 0 \\ 0 & 1/\sum (x_i - \bar{x}.)^2 \end{pmatrix}$$

Det i 'te diagonalled i hat-matricen er

$$\begin{pmatrix} 1 & (x_i - \bar{x}.) \end{pmatrix} (\mathbf{X}'\mathbf{X})^{-1} \begin{pmatrix} 1 \\ x_i - \bar{x}.) \end{pmatrix} = \frac{1}{k} + \frac{(x_i - \bar{x}.)^2}{\sum (x_j - \bar{x}.)^2}.$$

Det i 'te diagonalled udtrykker således den euklidiske afstand mellem punktet $(1, x_i)$ svarende til det i 'te datapunkt, og tyngdepunktet $(1/k, \bar{x})$ for alle datapunkter.

For en regressionsmodel med m forklarende variable finder man tilsvarende, at det i 'te diagonalled i hat-matricen er

$$h_{ii} = \frac{1}{k} + \frac{(x_{i1} - \bar{x}_{.1})^2}{\sum (x_{j1} - \bar{x}_{.1})^2} + \cdots + \frac{(x_{im} - \bar{x}_{.m})^2}{\sum (x_{jm} - \bar{x}_{.m})^2} . \quad (3.42)$$

altså netop den euklidiske afstand mellem vektoren af forklarende variable for den i 'te observation og tyngdepunktet $(1/k, \bar{x}_{.1}, \dots, \bar{x}_{.m})$ svarende til alle observationer.

Størrelsen (3.42) kaldes også Mahalanobis' afstand mellem datapunktet svarende til den i 'te observation og punktsværmen bestående af samtlige observationer.

Størrelsen $h_{ii} - 1/k$ er således et invariant mål for den kvadratiske afstand mellem $\mathbf{x}_i$ og tyngdepunktet for alle n punkter i koefficientrummet. ∇

Diagonalelementerne i hat-matricen udtrykker således hvor meget den tilsvarende række i $\mathbf{X}$ -matricen afviger fra samlingen af værdier af de forklarende variable.

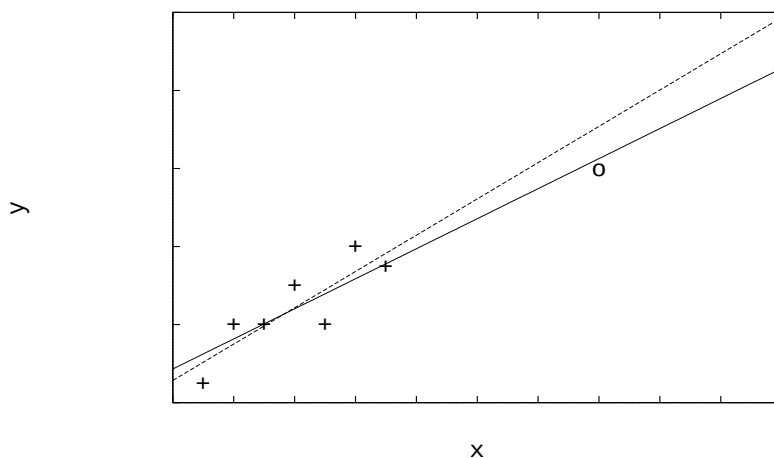
I engelsksproget litteratur siger man, at diagonalelementerne udtrykker den *leverage* (vægtstangseffekt), der er knyttet til den pågældende observation. Leverage er således en egenskab knyttet til modellen og "designet", dvs til værdierne af de forklarende variable; men som vi så i afsnit 3.4, er konsekvenserne af en stor leverage af betydning såvel for variansen på residualet, som for variansen på den fittede værdi.

Man skal have opmærksomheden rettet specielt mod observationer med stor leverage. Ikke fordi det nødvendigvis er en dårlig egenskab at en observation har stor leverage. Hvis et punkt med stor leverage i øvrigt er i overensstemmelse med resten af data, vil dette punkt blot tjene til at mindske usikkerheden på de estimerede koefficienter, se figur 3.6. Hvis en model er nogenlunde rimelig, og data er i overensstemmelse med modellen vil observationer med stor leverage være en fordel.

Observationer med stor leverage kan imidlertid være en indikation af at modellen ikke svarer til data. Det er velkendt, at man skal udvise stor varsomhed med at ekstrapolere en empirisk bestemt relation ud over det område i

“designrummet”, hvor den er tilpasset. En lineær model kan måske give en tilfredsstillende beskrivelse af data i et begrænset område af designrummet, men den behøver ikke nødvendigvis give en god beskrivelse uden for dette område.

Det er derfor ofte urealistisk at forvente, at man kan bestemme en lineær model, der giver en god beskrivelse i områder af designrummet, hvor der kun er få datapunkter. Datapunkter med stor leverage ligger netop i sådanne områder, og man kunne derfor formode, at sådanne observationer ikke beskrives særlig godt. Ofte ser man dog det modsatte fænomen. Modellen giver pæn tilpasning til datapunkter med stor leverage, mens tilpasningen til de øvrige datapunkter ikke er overvældende god, se f.eks. figur 3.7.



Figur 3.6: Illustration af et datapunkt $\circ$ med stor leverage

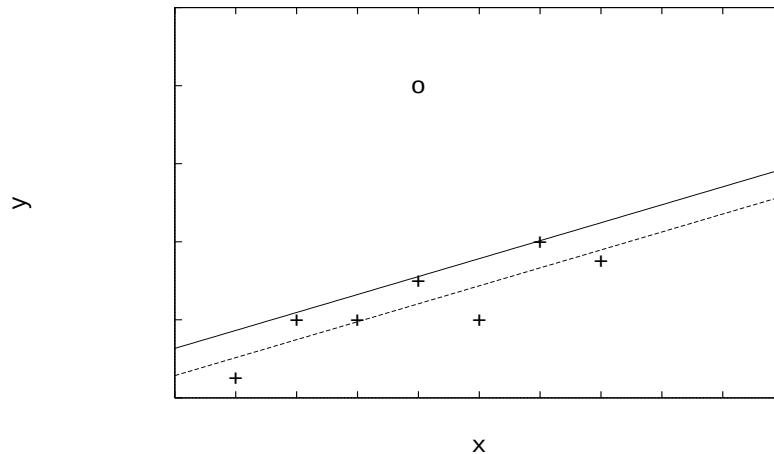
Den fuldt optrukne kurve angiver regressionslinien svarende til samtlige punkter. Den stiplede kurve viser regressionslinien estimeret under udeladelse af punktet $\circ$ med stor leverage.
(stor leverage, god konsistens, lille influens)

Belsley, Kuh og Welsch (1980) foreslår en afskæring på $2m/k$ for hat diago-

nalværdierne, hvor m er antallet af parametre i modellen, og k er antallet af observationer.

3.9.3 Kontrol af enkeltobservationers overensstemmelse, residual

Som vi har set, har de ustandardiserede residualer forskellig varians. Når man vil vurdere hvorvidt en enkelt observation afviger fra modellen, skal man derfor betragte de standardiserede eller de studentiserede residualer. Et standardiseret residual, der er større end 2-3, kan opfattes som tegn på en afvigende observation.



Figur 3.7: Illustration af et datapunkt $\circ$ med stort residual
Den fuldt optrukne kurve angiver regressionslinien svarende til samtlige punkter. Den stiplede kurve viser regressionslinien estimeret under udeladelse af punktet $\circ$ med et stort residual.
(lille leverage, ringe konsistens, lille influens)

3.9.4 Kontrol af enkeltobservationers indflydelse (influens)

Leverage

Vi har allerede diskuteret “leverage” som et udtryk for datapunktets afstand fra de øvrige punkter i designrummet og herved også som et sammenfattende udtryk for den vægt, der tillægges det pågældende datapunkt i den samlede estimation.

Man kan imidlertid også have interesse i at vurdere nogle mere specifikke mål for de enkelte datapunkters indflydelse på de enkelte komponenter i estimationsproblemet.

Den enkleste måde, hvorpå man kan måle indflydelsen fra et datapunkt er ved at sammenligne de estimer man får, når det pågældende datapunkt er medtaget i analysen og når det ikke er med.

I litteraturen er der foreslået en lang række mål til vurdering af indflydelsen fra et datapunkt. Nedenfor skal vi gennemgå en række af disse mål.

Belsley, Kuh og Welsch [32] og Cook og Weisberg [33] indeholder en mere udførlig beskrivelse.

I nedenstående gennemgang vil et indeks i rund parentes “(*i*)” indikere, at observation *i* er udeladt af estimationen.

De forskellige influensmål svarende til udeladelse af den *i*'te observation udregnes lettest ved at supplere **X**-matricen med en ekstra søjle med nuller i alle positioner på nær den *i*'te række. Effekten af denne søjle er, at modellen udvides med et led, der blot tager sig af at estimere y_i . De øvrige søjler i **X** bruges da til at estimere en model for de øvrige observationer.

Cook's *D*

Cook's *D* måler ændringerne i parameterestimer, hvis man udelader den *i*'te observation.

Cook's *D* svarende til den *i*'te observation er defineret som

$$D_i = \frac{(\hat{\beta} - \hat{\beta}_{(i)})' \mathbf{i}_\beta (\hat{\beta} - \hat{\beta}_{(i)})}{\hat{\sigma}^2 m} \quad (3.43)$$

hvor den forventede information $\mathbf{i}_\beta$ er

$$\mathbf{i}_\beta = \mathbf{X}' \Sigma^{-1} \mathbf{X}$$

Man ser, at Cook's D kan skrives

$$D_i = \frac{\|\hat{\boldsymbol{\eta}} - \hat{\boldsymbol{\eta}}_{(i)}\|_{\Sigma}^2}{\hat{\sigma}^2 m}$$

hvor $\hat{\boldsymbol{\eta}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ og $\hat{\boldsymbol{\eta}}_{(i)} = \mathbf{X}\hat{\boldsymbol{\beta}}_{(i)}$. Cook's D måler således forskellen i den lineære prædikator, når den i 'te observation fjernes. Såfremt der indgår en estimeret dispersionsparameter, er Cook's D skaleret med dette estimat (i dette estimat indgår dog den i 'te observation).

Cook (1977) foreslog at man sammenligner D_i med fraktilerne i en $F(m, k - m)$ -fordeling. En grov afskæring er ved $4/(k - m)$

Dffit og Dffits

Forskellen i fittet værdi, henholdsvis med og uden den i 'te observation kaldes $Dffit$.

$$F_i = \frac{\hat{\mu}_i - \hat{\mu}_{(i)}}{\sqrt{\hat{\sigma}^2 h_{ii}}}$$

Den tilsvarende skalerede værdi kaldes $Dffits$.

For den i 'te observation har man

$$F_i^* = \frac{\hat{\mu}_i - \hat{\mu}_{(i)}}{\sqrt{\hat{\sigma}_{(i)}^2 h_{ii}}} \quad (3.44)$$

hvor h_{ii} angiver diagonalelementet i hat-matricen.

Sædvanligvis bruger man en afskæring på 2 for DFFITS, En antalsjusteret afskæring er $\pm 2\sqrt{m/(k - m)}$

Dfbetas

Den normaliserede forskel på estimatet af β_j henholdsvis med og uden den i 'te observation kaldes $Dfbetas$.

For den i 'te observation har man

$$B_{j,i} = \frac{\hat{\beta}_j - \hat{\beta}_{j(i)}}{\sqrt{\hat{\sigma}_{(i)}^2 [\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}]_{jj}^{-1}}}, \quad (3.45)$$

hvor

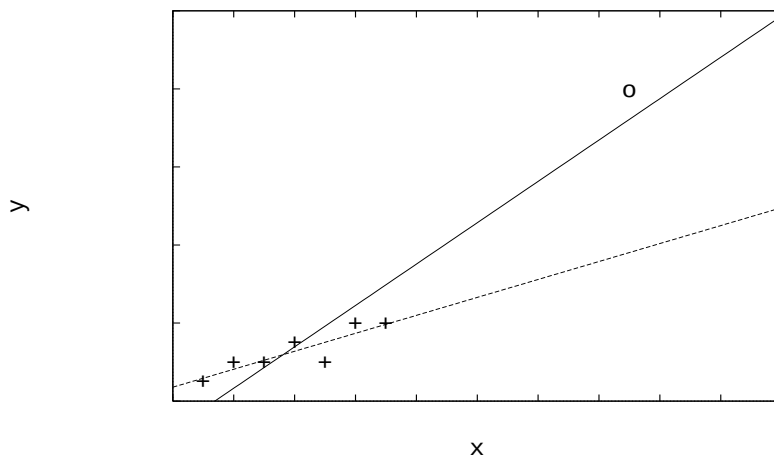
$$\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X} = \mathbf{I}_\beta$$

angiver informationsmatricen (med hensyn til β).

Man får en hel matrix af $B_{j,i}$ værdier; for hver observation får man en værdi for hver af de uafhængige variable.

Værdier af $B_{j,i}$ større end 2 indikerer observationer, der er influentielle i estimationen af en β_j .

En anbefalet afskæringsværdi er $\pm 2/\sqrt{k}$ hvor k angiver antallet af observationer.



Figur 3.8: Illustration af et datapunkt $\circ$ med stor influens

Den fuldt optrukne kurve angiver regressionslinien svarende til samtlige punkter. Den stiplede kurve viser regressionslinien estimeret under udeladelse af punktet $\circ$ med stor influens.
(stor leverage, ringe konsistens, stor influens)

3.10 Ortogonalitet, leverage, Lille eksempel og øvelse

Øvelse

Prøv at indtaste observationerne svarende til Ramus-eksemplet, afsnit 3.5 (både med rigtige aldre og med centrerede aldre), og udfør regressionsanalysen mht alder og mht centreret alder, kig på parameter estimates, på estimeret covariansmatrix, predicted, hat diagonal, residual, i de to tilfælde.

Hvornår er elementerne i hat-matricen store ?

3.11 Indskud om Yates' algoritme:

Når modellen kun indeholder faktorvariable med to niveauer, kan man opnå en ortogonal parametrisering ved at kode faktorniveauerne som -1 og 1 (type "Interval", dvs kontinuerte variable). Det er nu forholdsvis enkelt at danne søjlerne svarende til vekselvirkningsleddene ved multiplikation af søjlerne for de enkelte faktorer etc.

Denne fremgangsmåde sikrer en ortogonal dekomposition af den fulde model i enkeltbidrag med *uafhængige estimater* for de enkelte led. Leddene kan derfor testes hver for sig, dvs type III bidrag er de samme som type I bidrag. Led med forsvindende bidrag kan fjernes, og deres bidrag indgå i residualkvadratavgivelsessummen ved simpel addition. Da enkeltleddene kan fortolkes som meningsfulde kontraster, kan man i disse sammenhænge godt opleve modeller med indgående vekselvirkninger - uden at de tilsvarende hovedvirkninger indgår i modellen.

Til illustration af princippet beskrives nedenfor dekompositionen svarende til et fuldt faktorielt forsøg med de tre faktorer T, G, B (top, gear og bund for et husholdningsapparat) med to niveauer for hver faktor:

$$\mathbf{X} = \begin{bmatrix} 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & -1 & 1 & -1 & -1 & -1 \\ 1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 \\ 1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

med

$$\boldsymbol{\beta} = \begin{bmatrix} I \\ T \\ G \\ B \\ TG \\ TB \\ GB \\ TGB \end{bmatrix}$$

Hvis faktorerne betegnes x_1, x_2 og x_3 bliver parametervektoren

$$\boldsymbol{\beta} = \begin{bmatrix} I \\ x_1 \\ x_2 \\ x_3 \\ x_1x_2 \\ x_1x_3 \\ x_2x_3 \\ x_1x_2x_3 \end{bmatrix}$$

$$\mathbf{X7} := \begin{bmatrix} 1 & -1 & -1 & -1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 \\ 1 & 1 & 1 & -1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$\mathbf{X}' := \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\ -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \end{bmatrix}$$

$$\mathbf{X}'\mathbf{X} := \begin{bmatrix} 8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 8 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 8 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 8 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 8 \end{bmatrix}$$

$$(\mathbf{X}'\mathbf{X})^{-1} := \begin{bmatrix} \frac{1}{8} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{8} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{8} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{8} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{8} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{8} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{8} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{8} \end{bmatrix}$$

Lad observationsvektoren være

$$\mathbf{y} := \begin{bmatrix} y_{111} \\ y_{211} \\ y_{121} \\ y_{221} \\ y_{112} \\ y_{212} \\ y_{122} \\ y_{222} \end{bmatrix}$$

$$\mathbf{X}'\mathbf{y} := \begin{bmatrix} y_{111} + y_{211} + y_{121} + y_{221} + y_{112} + y_{212} + y_{122} + y_{222} \\ -y_{111} + y_{211} - y_{121} + y_{221} - y_{112} + y_{212} - y_{122} + y_{222} \\ -y_{111} - y_{211} + y_{121} + y_{221} - y_{112} - y_{212} + y_{122} + y_{222} \\ -y_{111} - y_{211} - y_{121} - y_{221} + y_{112} + y_{212} + y_{122} + y_{222} \\ y_{111} - y_{211} - y_{121} + y_{221} + y_{112} - y_{212} - y_{122} + y_{222} \\ y_{111} - y_{211} + y_{121} - y_{221} - y_{112} + y_{212} - y_{122} + y_{222} \\ y_{111} + y_{211} - y_{121} - y_{221} - y_{112} - y_{212} + y_{122} + y_{222} \end{bmatrix}$$

Man får da effektestimaterne

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} I \\ T \\ G \\ B \\ TG \\ TB \\ GB \\ TGB \end{bmatrix} = \frac{1}{8} \begin{bmatrix} y_{111} + y_{211} + y_{121} + y_{221} + y_{112} + y_{212} + y_{122} + y_{222} \\ -y_{111} + y_{211} - y_{121} + y_{221} - y_{112} + y_{212} - y_{122} + y_{222} \\ -y_{111} - y_{211} + y_{121} + y_{221} - y_{112} - y_{212} + y_{122} + y_{222} \\ -y_{111} - y_{211} - y_{121} - y_{221} + y_{112} + y_{212} + y_{122} + y_{222} \\ y_{111} - y_{211} - y_{121} + y_{221} + y_{112} - y_{212} - y_{122} + y_{222} \\ y_{111} - y_{211} + y_{121} - y_{221} - y_{112} + y_{212} - y_{122} + y_{222} \\ y_{111} + y_{211} - y_{121} - y_{221} - y_{112} - y_{212} + y_{122} + y_{222} \\ -y_{111} + y_{211} + y_{121} - y_{221} + y_{112} - y_{212} - y_{122} + y_{222} \end{bmatrix}$$

med middelværdier

$$\beta = \begin{bmatrix} I \\ T \\ G \\ B \\ TG \\ TB \\ GB \\ TGB \end{bmatrix} = \frac{1}{8} \begin{bmatrix} \mu_{111} + \mu_{211} + \mu_{121} + \mu_{221} + \mu_{112} + \mu_{212} + \mu_{122} + \mu_{222} \\ -\mu_{111} + \mu_{211} - \mu_{121} + \mu_{221} - \mu_{112} + \mu_{212} - \mu_{122} + \mu_{222} \\ -\mu_{111} - \mu_{211} + \mu_{121} + \mu_{221} - \mu_{112} - \mu_{212} + \mu_{122} + \mu_{222} \\ -\mu_{111} - \mu_{211} - \mu_{121} - \mu_{221} + \mu_{112} + \mu_{212} + \mu_{122} + \mu_{222} \\ \mu_{111} - \mu_{211} - \mu_{121} + \mu_{221} + \mu_{112} - \mu_{212} - \mu_{122} + \mu_{222} \\ \mu_{111} - \mu_{211} + \mu_{121} - \mu_{221} - \mu_{112} + \mu_{212} - \mu_{122} + \mu_{222} \\ \mu_{111} + \mu_{211} - \mu_{121} - \mu_{221} - \mu_{112} - \mu_{212} + \mu_{122} + \mu_{222} \\ -\mu_{111} + \mu_{211} + \mu_{121} - \mu_{221} + \mu_{112} - \mu_{212} - \mu_{122} + \mu_{222} \end{bmatrix} \quad (3.46)$$

$$\beta = \begin{bmatrix} I \\ T \\ G \\ B \\ TG \\ TB \\ GB \\ TGB \end{bmatrix} = \frac{1}{8} \begin{bmatrix} \mu_{...} \\ \mu_{2..} - \mu_{1..} \\ \mu_{.2.} - \mu_{.1.} \\ \mu_{..2} - \mu_{..1} \\ \mu_{11.} + \mu_{22.} - (\mu_{21.} + \mu_{12.}) \\ \mu_{1.1} + \mu_{2.2} - (\mu_{2.1} + \mu_{1.2}) \\ \mu_{.11} + \mu_{.22} - (\mu_{.21} + \mu_{.12}) \\ -\mu_{111} + \mu_{211} + \mu_{121} - \mu_{221} + \mu_{112} - \mu_{212} - \mu_{122} + \mu_{222} \end{bmatrix}$$

altså punkter på “terningen” $(-1, 1)^3$