Note til

Lineære Normale Modeller

Bo Rosbjerg

2. marts 2009

Tegninger udført af Henrik Vie Christensen

Indhold

1 En simpel lineær normal model

1.1 Modelbestemmelse

Lad $y_1, \ldots, y_n$ være observationer i en normalfordelt population med middelværdi μ og varians σ^2. Observationerne er realisationer af stokastiske variable

$$Y_i \sim N(\mu, \sigma^2) \ , \quad i = 1, \ldots, n \ , \quad \text{uafhængige.}$$

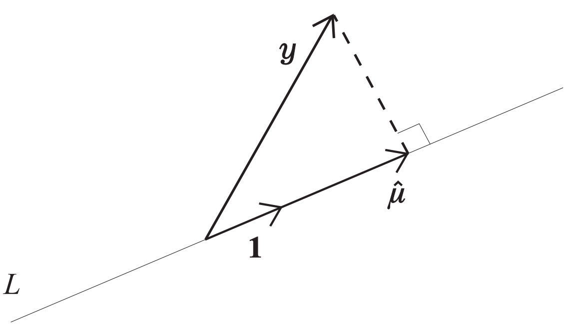
Vi kan samle observationerne i en n-dimensional vektor $\boldsymbol{y} = (y_1, \ldots, y_n) \in R^n$. Tilsvarende for middelværdierne, dvs. $\boldsymbol{\mu} = \mu(1, \ldots, 1) = \mu\mathbf{1}$. Det ses, at $\boldsymbol{\mu} \in L = \mathrm{sp}\,\{\mathbf{1}\}$, og at L er et underrum af R^n.

Modellen kaldes lineær, idet et vektorrum også betegnes som et lineært rum. I almindelighed gælder, at $-\infty < \mu < \infty$, dvs. at parameterrummet er $R \times R_+$

Modellen har to parametre, μ og σ^2, som skal estimeres.

1.2 Mindste kvadraters estimat

Lad $\hat{\boldsymbol{\mu}}$ betegne den ortogonale projektion af $\boldsymbol{y}$ på L, dvs. $\hat{\boldsymbol{\mu}} = P\boldsymbol{y} = \frac{\boldsymbol{y}\cdot\mathbf{1}}{\|\mathbf{1}\|^2}\mathbf{1} = \bar{y}\mathbf{1}$, hvor P er projektionsmatrix for ortogonalprojektionen på L. Ortogonalprojektionen på et underrum er en lineær transformation og kan derfor repræsenteres ved en matrix P. Her bliver $P = \frac{1}{n}\mathbf{1}\mathbf{1}^\top$, jf. opgave 1.



Figur 1: Ortogonalprojektion på et underrum.

Ved mindste kvadraters estimat forstås det valg af den/de ukendte parameter/-tre, som minimerer kvadratsummen af afstandene mellem de observerede værdier og de tilsvarende estimerede. Her er det altså størrelsen, $\|\boldsymbol{y} - \boldsymbol{\mu}\|^2$ som skal minimeres.

Den ovenfor indførte vektor $\hat{\boldsymbol{\mu}}$ ses let at være MK-estimatet for $\boldsymbol{\mu}$, idet der gælder

$$\|\boldsymbol{y} - \boldsymbol{\mu}\|^2 \geq \|\boldsymbol{y} - \hat{\boldsymbol{\mu}}\|^2 \quad \text{for alle } \boldsymbol{\mu} \in L \ ,$$

og

$$\|\boldsymbol{y} - \boldsymbol{\mu}\|^2 = \|\boldsymbol{y} - \hat{\boldsymbol{\mu}}\|^2 \Leftrightarrow \boldsymbol{\mu} = \hat{\boldsymbol{\mu}} \ ,$$

jf. opgave 2. $\bar{y}$ er altså MK-estimat for μ. Vi skriver $\hat{\mu} = \bar{y}$.

Bemærk, at $\bar{Y} \sim N(\mu, \frac{\sigma^2}{n})$.

1.3 Residualer

Forskellen mellem en observeret værdi og den tilsvarende estimerede værdi kaldes residualet:

$$r_i = y_i - \hat{y} = y_i - \hat{\mu} = y_i - \bar{y} \ , \quad i = 1, \ldots, n \ .$$

Bemærk, at både $\bar{y}$ og $y_i - \bar{y}$ er linearkombinationer af y_i'erne. Med resultater fra afsnit 3.3 gælder derfor, at de stokastiske variable $\bar{Y}$ og $Y_i - \bar{Y}$ begge er normalfordelte, og endvidere at de er uafhængige, idet

$$\begin{aligned}
\mathrm{Cov}\left(\bar{Y}, Y_i - \bar{Y}\right) &= \mathrm{Cov}\left(\frac{1}{n}\sum_j Y_j, Y_i\right) - \mathrm{Cov}\left(\bar{Y}, \bar{Y}\right) \\
&= \frac{1}{n}\sum_j \mathrm{Cov}\left(Y_j, Y_i\right) - \mathrm{Var}\left(\bar{Y}\right) \\
&= \frac{1}{n}(\mathrm{Var}\left(Y_i\right) + (n-1)0) - \frac{\sigma^2}{n} \\
&= \frac{\sigma^2}{n} - \frac{\sigma^2}{n} = 0 \ .
\end{aligned}$$

Bestemmelse af $\mathrm{Var}\left(R_i\right) = \mathrm{Var}\left(Y_i - \bar{Y}\right)$:

$$\begin{aligned}
Y_i = (Y_i - \bar{Y}) + \bar{Y} \quad &\Rightarrow \quad \mathrm{Var}\left(Y_i\right) = \mathrm{Var}\left(Y_i - \bar{Y}\right) + \mathrm{Var}\left(\bar{Y}\right) \\
&\Leftrightarrow \quad \sigma^2 = \mathrm{Var}\left(Y_i - \bar{Y}\right) + \frac{\sigma^2}{n} \\
&\Leftrightarrow \quad \mathrm{Var}\left(Y_i - \bar{Y}\right) = \sigma^2\left(1 - \frac{1}{n}\right) \ .
\end{aligned}$$

For residualet som stokastisk variabel gælder åbenbart, idet bestemmelsen af $\mathrm{E}\left[R_i\right]$ er triviel, at

$$R_i \sim N\left(0, \sigma^2\left(1 - \frac{1}{n}\right)\right) \ , \quad i = 1, \ldots, n \ .$$

Bemærk, at de enkelte residualer *ikke* er uafhængige; der gælder, at

$$\mathrm{Cov}\left(R_i, R_j\right) = -\frac{1}{n}\sigma^2 \ , \quad i \neq j \ ,$$

jf. opgave 3.

1.4 Residualkvadratsummen

For residualkvadratsummen gælder, at

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

jf. opgave 4.

Betragter vi de tilsvarende stokastiske variable, ser vi, at

$$\begin{aligned} \mathrm{E}\left[\sum_i \left(Y_i - \bar{Y}\right)^2\right] &= \sum_i \mathrm{E}\left[Y_i - \mu\right]^2 - n\mathrm{E}\left[\bar{Y} - \mu\right]^2 \\ &= n\sigma^2 - n\frac{\sigma^2}{n} \ , \qquad \text{jf. definitionen af varians} \\ &= (n-1)\,\sigma^2 \ . \end{aligned}$$

Der gælder åbenbart, at størrelsen $\frac{1}{n-1}\sum_i \left(y_i - \bar{y}\right)^2$ er et centralt estimat for σ^2. Størrelsen betegnes normalt s^2.

Som stokastisk variabel er S^2 uafhængig af $\bar{Y}$. Dette resultat vises sidst i afsnit 3.3.

Med henblik på at bestemme fordelingen af S^2 betragter vi igen

$$\begin{aligned} \sum_i (Y_i - \bar{Y})^2 &= \sum_i (Y_i - \mu)^2 - n(\bar{Y} - \mu)^2 \\ \Leftrightarrow \quad \sum_i \left(\frac{Y_i - \mu}{\sigma}\right)^2 &= \frac{1}{\sigma^2}\sum_i (Y_i - \bar{Y})^2 + \left(\frac{\bar{Y} - \mu}{\frac{\sigma}{\sqrt{n}}}\right)^2 \ . \end{aligned}$$

De to led på højresiden er uafhængige, da $\bar{Y}$ og S^2 er uafhængige. Endvidere har vi, at

$$\sum_i \left(\frac{Y_i - \mu}{\sigma}\right)^2 \sim \chi^2(n) \quad \text{og} \quad \left(\frac{\bar{Y} - \mu}{\frac{\sigma}{\sqrt{n}}}\right)^2 \sim \chi^2(1) \ ,$$

jf. definition af χ^2-fordelingen, se afsnit 1.13.

For de momentfrembringende funktioner[1] gælder derfor, at

$$\begin{aligned} & M_{\sum_i \left(\frac{Y_i-\mu}{\sigma}\right)^2}(t) = M_{\frac{1}{\sigma^2}\sum_i (Y_i - \bar{Y})^2}(t) M_{\left(\frac{\bar{Y}-\mu}{\frac{\sigma}{\sqrt{n}}}\right)^2}(t) \\ \Leftrightarrow \quad & (1-2t)^{-\frac{n}{2}} = M_{\frac{1}{\sigma^2}\sum_i (Y_i - \bar{Y})^2}(t)(1-2t)^{-\frac{1}{2}} \\ \Leftrightarrow \quad & M_{\frac{1}{\sigma^2}\sum_i (Y_i - \bar{Y})^2}(t) = (1-2t)^{-\frac{n-1}{2}} \\ \Leftrightarrow \quad & \tfrac{1}{\sigma^2}\textstyle\sum_i (Y_i - \bar{Y})^2 \sim \chi^2(n-1) \ . \end{aligned}$$

[1] Den momentfrembringende funktion for en stokastisk variabel Y defineres som $M_Y(t) = \mathrm{E}\, e^{tY}$, hvor t er en reel variabel. Bemærk, at $\mathrm{E}\left[Y^j\right] = M_Y^{(j)}(0)$. For uafhængige stokastiske variable X og Y ses, at $M_{X+Y}(t) = M_X(t)M_Y(t)$. Vedrørende entydighed: Når to stokastiske variabler har samme momentfrembringende funktion, har de også samme sandsynlighedsfordeling.

Bemærk, at $\frac{1}{\sigma^2}\sum_i (Y_i - \bar{Y})^2 = \frac{(n-1)S^2}{\sigma^2}$, hvorefter vi har, at

$$\begin{aligned}
S^2 &\sim \frac{\sigma^2}{n-1}\chi^2(n-1) \ , \\
\mathrm{E}\left[S^2\right] &= \frac{\sigma^2}{n-1}(n-1) = \sigma^2 \qquad \text{(regnekontrol)}, \\
\mathrm{Var}\left(S^2\right) &= \left(\frac{\sigma^2}{n-1}\right)^2 2(n-1) = \frac{2\sigma^4}{n-1} \ .
\end{aligned}$$

1.5 Sammenfatning

$$\begin{array}{ll}
\hat{\mu} = \bar{y} \ , & \bar{Y} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right) \ , \\
r_i = y_i - \bar{y}, \quad i = 1, \ldots, n \ , & R_i \sim N\left(0, \sigma^2\left(1 - \dfrac{1}{n}\right)\right), \quad i = 1, \ldots, n \ , \\
 & \mathrm{Cov}\left(R_i, R_j\right) = -\dfrac{1}{n}\sigma^2 \ , \quad i \neq j \ , \\
s^2 = \dfrac{1}{n-1}\displaystyle\sum_i (y_i - \bar{y})^2 \ , & S^2 \sim \dfrac{\sigma^2}{n-1}\chi^2(n-1) \ , \\
 & \bar{Y} \text{ og } S^2 \text{ uafhængige.}
\end{array}$$

1.6 Maksimum likelihood estimation

Likelihoodfunktionen for (μ, σ^2) er defineret ved

$$L(\mu, \sigma^2; y_1, \ldots, y_n) = f_{\boldsymbol{Y}}(y_1, \ldots, y_n; \mu, \sigma^2),$$

hvor $f_{\boldsymbol{Y}}$ er den simultane tæthedsfunktion for $Y_1, \ldots, Y_n$, dvs.

$$\begin{aligned}
L(\mu, \sigma^2) &= \prod_i \frac{1}{\sqrt{2\pi}\sigma} \ \mathrm{e}^{-\frac{(y_i-\mu)^2}{2\sigma^2}} \\
&= (2\pi\sigma^2)^{-\frac{n}{2}} \mathrm{e}^{-\frac{1}{2\sigma^2}\sum_i (y_i-\mu)^2} \\
&= (2\pi\sigma^2)^{-\frac{n}{2}} \mathrm{e}^{-\frac{1}{2\sigma^2}\|\boldsymbol{y}-\mu\mathbf{1}\|^2},
\end{aligned}$$

og dermed loglikelihoodfunktionen

$$l(\mu, \sigma^2) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\|\boldsymbol{y} - \mu\mathbf{1}\|^2$$

Den værdi i parameterrummet, som giver den største værdi af likelihoodfunktionen, opfattes som den, der bedst understøttes af de foreliggende observationer.

Bemærk, at L og l har supremum for samme værdi af (μ, σ^2), idet ln er er en monoton funktion.

Det ses umiddelbart, at

$$\forall \sigma^2 > 0 : \sup_{\mu \in R} l(\mu, \sigma^2) = l(\hat{\mu}, \sigma^2),$$

hvor $\hat{\mu} = \bar{y}$, jf. afsnit 1.2.

Funktionen $\check{l}(\sigma^2) = l(\hat{\mu}, \sigma^2)$ kaldes profilloglikelihoodfunktionen for σ^2. Udtrykket for $\check{l}$ bliver

$$\check{l}(\sigma^2) = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \|\boldsymbol{y} - \bar{y}\mathbf{1}\|^2.$$

Ved differentiation fås

$$\check{l}'(\sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \|\boldsymbol{y} - \bar{y}\mathbf{1}\|^2,$$

hvoraf

$$\check{l}'(\sigma^2) = 0 \quad \text{for} \quad \hat{\sigma^2} = \frac{1}{n} \|\boldsymbol{y} - \bar{y}\mathbf{1}\|^2.$$

At det er et maksimumspunkt, vi har bestemt, ses af

$$\check{l}''(\sigma^2)\Big|_{\sigma^2 = \hat{\sigma^2}} = \frac{n}{2(\sigma^2)^2} - \frac{1}{(\sigma^2)^3} \|\boldsymbol{y} - \bar{y}\mathbf{1}\|^2\Big|_{\sigma^2 = \hat{\sigma^2}} = \frac{n}{2(\hat{\sigma^2})^2} - \frac{n\hat{\sigma^2}}{(\hat{\sigma^2})^3} = -\frac{n}{2(\hat{\sigma^2})^2} < 0.$$

Den maksimale værdi af likelihoodfunktionen bliver

$$L(\hat{\mu}, \hat{\sigma}^2) = (2\pi\hat{\sigma^2})^{-\frac{n}{2}} \, \mathrm{e}^{-\frac{1}{2\hat{\sigma^2}} n\hat{\sigma^2}} = (2\pi e\hat{\sigma^2})^{-\frac{n}{2}}.$$

Det ses, at MK- og ML-estimaterne for μ er sammenfaldende. Som estimat for σ^2 benytter vi normalt ikke $\hat{\sigma^2}$, men det s^2 vi bestemte i forbindelse med MK-estimation, bl.a. for at opnå et centralt estimat. For sammenhængen mellem $\hat{\sigma^2}$ og s^2 gælder, at $s^2 = \frac{n}{n-1}\hat{\sigma^2}$.

1.7 Sufficient reduktion

Idet

$$\begin{aligned} f_{\boldsymbol{Y}}(y_1, \ldots, y_n; \mu, \sigma^2) &= (2\pi\sigma^2)^{-\frac{n}{2}} \, \mathrm{e}^{-\frac{1}{2\sigma^2} \sum_i (y_i - \mu)^2} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \, \mathrm{e}^{-\frac{1}{2\sigma^2} (\sum_i {y_i}^2 - 2n\mu\bar{y} + n\mu^2)} \end{aligned}$$

ses ved benyttelse af Neymans kriterium[2], at $(\bar{y}, \sum_i y_i^2)$ er en sufficient reduktion for (μ, σ^2), subsidiært kan benyttes $(\sum_i y_i, \sum_i y_i^2)$ eller $(\bar{y}, \sum_i (y_i - \bar{y})^2)$ eller andre varianter.

1.8 Modelkontrol

At observationerne stammer fra en normalfordelt population, kan eftervises grafisk ved et såkaldt normalfraktildiagram.

Lad $y_{(1)}, \ldots, y_{(n)}$ betegne observationerne ordnet efter størrelse.

Den kumulerede relative frekvens svarende til den i'te observation efter størrelse er

$$f_i = \frac{i}{n} \ , \quad i = 1, \ldots, n \ .$$

Lad $F_n(y)$ betegne den empiriske fordelingsfunktion, dvs.

$$F_n(y) = \frac{1}{n} \sum_i I_i(y_{(i)} \leq y) \ ,$$

hvor I_i er indikatorvariabel for hændelsen $y_{(i)} \leq y$. For ethvert fastholdt y må der gælde, at $nF_n(y) \sim b(n, p)$, hvor $p = P(Y_i \leq y) = F(y)$; F er fordelingsfunktion for observationerne. De store tals lov viser, at $F_n(y) \to F(y)$ for $n \to \infty$. For at få en hurtigere konvergens modificeres f_i ofte ved benyttelse af en konstant a til

$$f_i = \frac{i - a}{n + 1 - 2a} \ , \qquad 0 \leq a \leq 1 \ , \qquad i = 1, \ldots, n \ ,$$

hvor de mest almindelige valg af a er $\frac{3}{8}$, når $n \leq 10$, eller $\frac{1}{2}$, når $n > 10$.

Når observationerne er normalfordelte, må der gælde, at

$$f_i \approx \Phi\left(\frac{y_{(i)} - \mu}{\sigma}\right) \ , \quad i = 1, \ldots, n \ ,$$

som er ækvivalent med

$$\Phi^{-1}(f_i) \approx \frac{y_{(i)} - \mu}{\sigma} \ , \quad i = 1, \ldots, n \ .$$

Sætter vi $\Phi^{-1}(f_i) = u_i$, $i = 1, \ldots, n$, skal punkterne $(y_{(i)}, u_i)$, $i = 1, \ldots, n$, tilnærmelsesvis ligge på en ret linie med hældningskoefficient $\frac{1}{\sigma}$. Linien afskærer stykket $\frac{\mu}{\sigma}$ på u-aksen.

[2]Betragt en model med parameter θ (θ kan være flerdimensional). Neymans kriterium siger, at stikprøvefunktionen $t(y_1, \ldots, y_n)$ er sufficient for θ, når og kun når den simultane tæthedsfunktion for $Y_1, \ldots, Y_n$ kan faktoriseres som

$$f(y_1, \ldots, y_n; \theta) = h(y_1, \ldots, y_n) g(t(y_1, \ldots, y_n); \theta).$$

Bemærk, at h ikke afhænger af θ, og at g kun afhænger af θ gennem $t(y_1, \ldots, y_n)$.

På det såkaldte sandsynlighedspapir svarer inddelingen på den ene akse til transformationen Φ^{-1}, således at f_i'erne kan afsættes direkte.

Normalfraktildiagrammet kan benyttes til grafisk estimation af μ svarende til $u = 0$, og σ fx som halvdelen af afstanden mellem y svarende til $u = 2$ og det estimerede μ.

1.9 Konfidensintervaller

I tilknytning til et parameterestimat er det ofte ønskeligt, at kunne angive et skøn over, hvor præcist det pågældende estimat er, eller sagt på en anden måde, hvor stor tiltro vi tør have til den beregnede værdi. Her kommer de såkaldte konfidensintervaller ind i billedet.

Lad os gå direkte i gang med at konstruere et konfidensinterval for parameteren μ i vores model.

Som vi har benyttet flere gange i det foregående, gælder der, at $\bar{Y} \sim N(\mu, \frac{\sigma^2}{n})$ og $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$, $\bar{Y}$ og S^2 uafhængige, og dermed at

$$\frac{\bar{Y}-\mu}{\frac{S}{\sqrt{n}}} = \frac{\frac{\bar{Y}-\mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{\frac{(n-1)S^2}{\sigma^2}}{n-1}}} \sim t(n-1) \ ,$$

jf. definition af t-fordelingen, se afsnit 1.14.

Følgende sandsynlighedsudsagn om den t-fordelte variabel må være sandt:

$$P\left(-t_{1-\frac{\alpha}{2}}(n-1) \leq \frac{\bar{Y}-\mu}{\frac{S}{\sqrt{n}}} \leq t_{1-\frac{\alpha}{2}}(n-1)\right) = 1-\alpha \ ,$$

hvor $t_{1-\frac{\alpha}{2}}(n-1)$ betegner $(1-\frac{\alpha}{2})$-fraktilen[3] i t-fordelingen med $n-1$ frihedsgrader.

Sandsynlighedsudsagnet kan omformes til

$$P\left(\bar{Y} - t_{1-\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}} \leq \mu \leq \bar{Y} + t_{1-\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}}\right) = 1-\alpha \ .$$

Erstatter vi de stokastiske variable $\bar{Y}$ og S^2 med deres realiserede værdier $\bar{y}$ og s^2, får vi et udsagn, som er sandt eller falsk:

$$\bar{y} - t_{1-\frac{\alpha}{2}}(n-1)\frac{s}{\sqrt{n}} \leq \mu \leq \bar{y} + t_{1-\frac{\alpha}{2}}(n-1)\frac{s}{\sqrt{n}} \ .$$

[3]Ved p-fraktilen i en fordeling med fordelingsfunktion $F(x)$ forstås $F^{-1}(p)$, $0 < p < 1$. p-fraktilen skrives ofte x_p, altså $x_p = F^{-1}(p)$.

Vi er ude af stand til at afgøre sandhedsværdien af udsagnet, men vi kan vælge at have en grad af tiltro til udsagnet svarende til sandsynligheden i det bagved liggende sandsynlighedsudsagn. Vi siger, at vi har konstrueret et konfidensinterval for parameteren μ med konfidensgrad $1-\alpha$.

Konfidensintervallet for μ skrives ofte på en lidt upræcis, men praktisk form:

$$\mu = \bar{y} \pm t_{1-\frac{\alpha}{2}}(n-1)\frac{s}{\sqrt{n}} \ .$$

Betragter vi igen den stokastiske variabel $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$, kan vi også konstruere et konfidensinterval for σ^2, idet

$$P\left(\chi^2_{\frac{\alpha}{2}}(n-1) \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{1-\frac{\alpha}{2}}(n-1)\right) = 1-\alpha \ ,$$

som omformes til

$$P\left(\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}\right) = 1-\alpha \ .$$

Konfidensintervallet for parameteren σ^2 med konfidensgrad $1-\alpha$ bliver da

$$\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}(n-1)} \ .$$

1.10 Hypotesetest

Et almindeligt udgangspunkt for at konstruere et test er at benytte den såkaldte kvotientteststørrelse, som har likelihoodfunktionens maksimale størrelse i nævneren og likelihoodfunktionens maksimale størrelse under nulhypotesen i tælleren. Store værdier af kvotientteststørrelsen understøtter derfor nulhypotesen, mens små værdier er kritiske for denne.

Her vil vi ved anvendelse af principperne for kvotienttest udvikle det såkaldte t-test for hypotesen $H_0 : \mu = \mu_0$. Som alternativ hypotese er der tre forskellige muligheder $H_1 : \mu < \mu_0$, $H_1 : \mu \neq \mu_0$ eller $H_1 : \mu > \mu_0$. Testet med det første eller det sidste alternativ kaldes ensidet, henholdsvis nedad eller opad, mens det midterste alternativ svarer til et tosidet test.

For at kunne udlede kvotienttestet for $H_0 : \mu = \mu_0$ må vi først estimere i den modificerede model. Det er kun parameteren σ^2, der skal estimeres. Maksimum likelihood estimation fører umiddelbart til resultatet $\hat{\hat{\sigma}}^2 = \frac{1}{n}\sum_i(y_i-\mu_0)^2$ og likelihoodfunktionens maksimale værdi bliver $L(\hat{\hat{\sigma}}^2) = (2\pi e\hat{\hat{\sigma}}^2)^{-\frac{n}{2}}$, jf. opgave 8.

Kvotientteststørrelsen opstilles:

$$\begin{aligned} q(y_1,\ldots,y_n) &= \frac{L(\hat{\hat{\sigma^2}})}{L(\hat{\mu},\hat{\sigma^2})} = \frac{(2\pi e\hat{\hat{\sigma^2}})^{-\frac{n}{2}}}{(2\pi e\hat{\sigma^2})^{-\frac{n}{2}}} = \left(\frac{\hat{\sigma^2}}{\hat{\hat{\sigma^2}}}\right)^{\frac{n}{2}} = \left(\frac{\frac{1}{n}\sum_i(y_i-\bar{y})^2}{\frac{1}{n}\sum_i(y_i-\mu_0)^2}\right)^{\frac{n}{2}} \\ &= \left(\frac{\sum_i(y_i-\bar{y})^2}{\sum_i(y_i-\bar{y})^2+n(\bar{y}-\mu_0)^2}\right)^{\frac{n}{2}} = \left(\frac{1}{1+\dfrac{n(\bar{y}-\mu_0)^2}{\sum_i(y_i-\bar{y})^2}}\right)^{\dfrac{n}{2}} . \end{aligned}$$

Størrelsen $(n-1)\frac{n(\bar{y}-\mu_0)^2}{\sum_i(y_i-\bar{y})^2}$ er en aftagende transformation af q, så store værdier af denne størrelse er kritiske for H_0.

Bemærk, at

$$(n-1)\frac{n(\bar{y}-\mu_0)^2}{\sum_i(y_i-\bar{y})^2} = (n-1)\frac{n(\bar{y}-\mu_0)^2}{(n-1)s^2} = \left(\frac{\dfrac{\bar{y}-\mu_0}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\dfrac{\frac{(n-1)s^2}{\sigma^2}}{(n-1)}}}\right)^2 = (t_{obs})^2 ,$$

hvor t_{obs} er realiseret værdi af en t-fordelt variabel med $n-1$ frihedsgrader.

Bemærk endvidere, at $(t_{obs})^2 = f_{obs}$, hvor f_{obs} er en realiseret værdi af en F-fordelt variabel med 1 tællerfrihedsgrad og $n-1$ nævnerfrihedsgrader, jf. definition af F-fordelingen, se afsnit 1.15. Vi kan altså vælge mellem at udføre et t-test eller et F-test. Vi vælger normalt t-testet af hensyn til simplere beregninger af testsandsynligheden ved ensidet test.

Testsandsynligheden svarende til de tre forskellige alternativer sættes til

$$\begin{array}{ll} P(T \leq t_{obs}) & \text{nedad ensidet test,} \\ 2P(T \geq |t_{obs}|) & \text{tosidet test,} \\ P(T \geq t_{obs}) & \text{opad ensidet test,} \end{array}$$

hvor $T \sim t(n-1)$ under H_0.

Det er også muligt, at konstruere et test for hypotesen $H_0 : \sigma^2 = \sigma_0^2$. Her er der igen tre muligheder for alternativ hypotese H_1. I den modificerede model under H_0 er det kun μ, der skal estimeres. ML-estimatet bliver $\hat{\hat{\mu}} = \sum_i(y_i-\bar{y})^2$ og $L(\hat{\hat{\mu}}) = (2\pi e^{\frac{\hat{\sigma^2}}{\sigma_0^2}}\sigma_0^2)^{-\frac{n}{2}}$, jf. opgave 11.

For kvotientteststørrelsen får vi

$$q(y_1,\ldots,y_n) = \frac{L(\hat{\hat{\mu}})}{L(\hat{\mu},\hat{\sigma^2})} = \frac{(2\pi e^{\frac{\hat{\sigma^2}}{\sigma_0^2}}\sigma_0^2)^{-\frac{n}{2}}}{(2\pi e\hat{\sigma^2})^{-\frac{n}{2}}} = \left(\frac{\hat{\sigma^2}}{\sigma_0^2}\; e^{1-\frac{\hat{\sigma^2}}{\sigma_0^2}}\right)^{\frac{n}{2}} .$$

Det ses, at $\frac{n\hat{\sigma^2}}{\sigma_0^2}$ er en aftagende transformation af q for $\frac{\hat{\sigma^2}}{\sigma_0^2} > 1$, voksende for $\frac{\hat{\sigma^2}}{\sigma_0^2} < 1$.

Bemærk $\frac{n\hat{\sigma^2}}{\sigma_0^2} = \frac{(n-1)s^2}{\sigma_0^2} = \chi^2_{obs}$, hvor χ^2_{obs} er en realiseret værdi af en χ^2-fordelt variabel med $n-1$ frihedsgrader.

Testsandsynligheden sættes til

$$\begin{array}{ll} P(X^2 < \chi^2_{obs}) & \text{nedad ensidet,} \\ 2\min\left\{P(X^2 < \chi^2_{obs}), P(X^2 > \chi^2_{obs})\right\} & \text{tosidet,} \\ P(X^2 > \chi^2_{obs}) & \text{opad ensidet,} \end{array}$$

hvor $X^2 \sim \chi^2(n-1)$ under H_0.

1.11 Signifikanstest

Ofte vælger man på forhånd at bestemme et kritisk område (forkastelsesområde) og dermed også et acceptområde for værdien af den betragtede teststørrelse. Der fastsættes et såkaldt signifikansniveau α, og ud fra dette bestemmes den/de kritiske værdi(er) som passende fraktiler i teststørrelsens fordeling under H_0. For fx et t-test bliver det kritiske område bestemt som

$$\begin{array}{ll} -\infty < t < t_\alpha & \text{nedad ensidet test,} \\ -\infty < t < -t_{1-\frac{\alpha}{2}} \text{ og } t_{1-\frac{\alpha}{2}} < t < \infty & \text{tosidet test,} \\ t_{1-\alpha} < t < \infty & \text{opad ensidet test.} \end{array}$$

Kommer t_{obs} til at ligge i det kritiske område, må vi forkaste H_0, og vi siger, at der er signifikans på niveau α. Mere udførligt opfatter vi den fremkomne værdi af t_{obs} som signifikant afvigende i forhold til det, vi med det givne signifikansniveau måtte forvente under H_0.

Kritiske områder for χ^2-testet:

$$\begin{array}{ll} 0 < \chi^2 < \chi^2_\alpha & \text{nedad ensidet,} \\ 0 < \chi^2 < \chi^2_{\frac{\alpha}{2}} \text{ og } \chi^2_{1-\frac{\alpha}{2}} < \chi^2 < \infty & \text{tosidet,} \\ \chi^2_{1-\alpha} < \chi^2 < \infty & \text{opad ensidet.} \end{array}$$

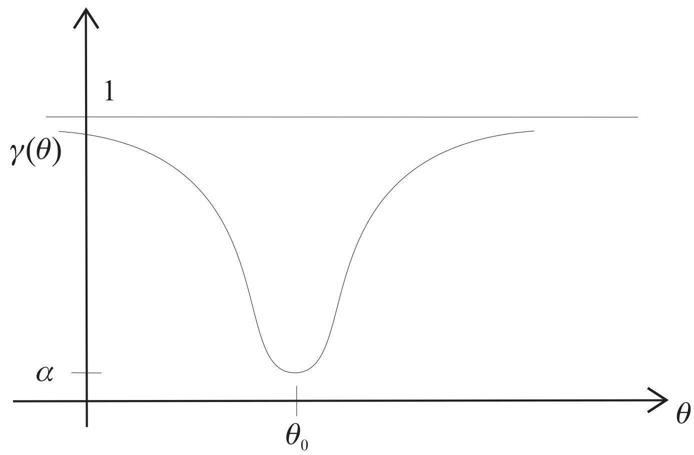
1.12 Fejlslutninger ved hypotesetest

Bemærk, at der er mulighed for to forskellige typer fejlslutning ved test af hypotesen $H_0 : \theta = \theta_0$, se skema:

	H_0 accepteres	H_0 forkastes
H_0 sand	Korrekt	Fejl af type I
H_0 falsk	Fejl af type II	Korrekt

Ved et signifikanstest er sandsynligheden for at begå en fejl af type I lig med det valgte signifikansniveau.

Testets evne til at forkaste en falsk hypotese vil normalt være en funktion af den parameter, testet handler om. Denne funktion betegnes testets styrkefunktion. Hvis $\beta(\theta)$ betegner sandsynligheden for at begå en fejl af type II, kan styrkefunktionen udtrykkes som $\gamma(\theta) = 1 - \beta(\theta)$. En styrkefunktion vil altid gå gennem punktet (θ_0, α), hvor α er signifikansniveauet.



Figur 2: Illustration af styrkefunktionen for testet $H_0 : \theta = \theta_0$, $H_1 : \theta \neq \theta_0$

Hvis vi stiller krav om, at testets styrke mindst skal være på værdien γ svarende til at givet θ, altså $\gamma(\theta) \geq \gamma$, kan dette normalt opfyldes ved en passende forøgelse af antallet af observationer.

1.13 χ^2-fordelinger

Med $U_i \sim N(0,1),\ i = 1, \ldots, k$ og $U_1, \ldots, U_k$ uafhængige, siger vi, at

$$V = {U_1}^2 + \ldots + {U_k}^2$$

er χ^2-fordelt med k frihedsgrader, og vi skriver $V \sim \chi^2(k)$.

V antager kun positive værdier. De almindeligt benyttede fraktilværdier er tabellagte, se fx Erlang S, der angiver 16 fraktilværdier for alle χ^2-fordelinger med frihedsgrader fra 1 til 100.

Fraktilværdier kan også bestemmes i R eller andre statistikpakker.

For $V \sim \chi^2(k)$ gælder

$$\mathrm{E}\,[V] = k\,, \qquad \mathrm{Var}\,(V) = 2k\,, \qquad M_V(t) = (1-2t)^{-\frac{k}{2}}\,, \quad t < \frac{1}{2}.$$

1.14 t-fordelinger

Med $U \sim N(0,1)$ og $V \sim \chi^2(k)$, U og V uafhængige, siger vi, at

$$T = \frac{U}{\sqrt{\frac{V}{k}}}$$

er t-fordelt med k frihedsgrader, og vi skriver $T \sim t(k)$.

t-fordelingens tæthedsfunktion er symmetrisk om $t = 0$.

For $k \to \infty$ konvergerer T i fordeling mod $U \sim N(0,1)$. $t(1)$ er identisk med Cauchy-fordelingen med $\theta = 0$.

Udvalgte fraktiler i de almindeligt benyttede t-fordelinger er tabellagte.

1.15 F-fordelinger

Med $V \sim \chi^2(k)$ og $W \sim \chi^2(m)$, V og W uafhængige, siger vi, at

$$F = \frac{\frac{V}{k}}{\frac{W}{m}} = \frac{mV}{kW}$$

er F-fordelt med k tællerfrihedsgrader og m nævnerfrihedsgrader, og vi skriver $F \sim F(k,m)$.

F-fordelte variable antager kun positive værdier.

For $m \to \infty$ konvergerer F i fordeling mod $\frac{1}{k}V$. Med $T \sim t(k)$ gælder, at $T^2 \sim F(1,k)$.

Også for de almindeligt benyttede F-fordelinger er udvalgte fraktiler tabellagte, dog kun for $p > \frac{1}{2}$. Fraktilværdier for $p < \frac{1}{2}$ kan, som det vises nedenfor, fås af

$$f_p(k,m) = \frac{1}{f_{1-p}(m,k)}\ .$$

I henhold til definitionen på F-fordeling har vi umiddelbart, at $\frac{1}{F} \sim F(m,k)$, hvilket udnyttes i følgende regninger, hvor F benyttes som betegnelse for både fordelingsfunktion og for den omhandlede stokastiske variabel:

$$
\begin{aligned}
& p = F_F(f_p(k,m)) = P(F \leq f_p(k,m)) = P\left(\frac{1}{F} \geq \frac{1}{f_p(k,m)}\right) \\
\Leftrightarrow \quad & 1-p = P\left(\frac{1}{F} < \frac{1}{f_p(k,m)}\right) = F_{\frac{1}{F}}\left(\frac{1}{f_p(k,m)}\right) \\
\Leftrightarrow \quad & f_{1-p}(m,k) = \frac{1}{f_p(k,m)} \ .
\end{aligned}
$$

1.16 Opgaver

1. Udled, at $P = \frac{1}{n}\mathbf{1}\mathbf{1}^\top$. Vink: Omskriv udtrykket $\bar{y}\mathbf{1}$ til $P\boldsymbol{y}$.

2. Eftervis, at $\|\boldsymbol{y}-\boldsymbol{\mu}\|^2 \geq \|\boldsymbol{y}-\hat{\boldsymbol{\mu}}\|^2$ med lighedstegn, når og kun når $\boldsymbol{\mu} = \hat{\boldsymbol{\mu}}$.

   Vink: $\|\boldsymbol{y}-\boldsymbol{\mu}\|^2 = \|\boldsymbol{y}-\hat{\boldsymbol{\mu}}+\hat{\boldsymbol{\mu}}-\boldsymbol{\mu}\|^2 = \ldots$

3. Eftervis, at $\mathrm{Cov}\,(R_i, R_j) = -\frac{1}{n}\sigma^2$, $i \neq j$. Vink: Udnyt, at kovariansoperatoren er bilineær.

4. Eftervis, at $\sum_i (y_i - \bar{y})^2 = \sum_i (y_i - \mu)^2 - n(\bar{y}-\mu)^2 = \sum_i y_i^2 - n\bar{y}^2$.

5. Vis, at $\sum_i {r_i}^2 = \boldsymbol{y}^\top (I-P)\boldsymbol{y}$, hvor $P = \frac{1}{n}\,\mathbf{1}\,\mathbf{1}^\top$. Sæt dernæst $\boldsymbol{u} = \boldsymbol{y} - \boldsymbol{\mu}$, og vis, at der også gælder, at $\sum_i {r_i}^2 = \boldsymbol{u}^\top (I-P)\boldsymbol{u}$.

6. Generer nogle forskellige normalfordelte datasæt, og optegn i hvert tilfælde et normalfraktildiagram.

7. Hvor stor er chancen i procent for, at et givet konfidensinterval omslutter den sande parameterværdi?

8. Bestem ML-estimatet for σ^2 i modellen $Y_i \sim N(\mu, \sigma^2)$, $i = 1, \ldots, n$, hvor μ er kendt. Angiv endvidere likelihoodfunktionens maksimale værdi.

9. Vis, at $T \sim t(m) \Rightarrow T^2 \sim F(1, m)$.

10. Vis, at $\{\mu_0 \mid H_0 : \mu = \mu_0 \text{ accepteres}\}$ ved tosidet signifikanstest på niveau α netop udgøres af konfidensintervallet for μ med konfidensgrad $1-\alpha$.

11. Bestem ML-estimatet for μ i modellen $Y_i \sim N(\mu, \sigma^2)$, $i = 1, \ldots, n$, hvor σ^2 er kendt. Angiv endvidere likelihoodfunktionens maksimale værdi.

12. Konstruer et konfidensinterval for parameteren μ i modellen i opgave 11.

    Vink: $\frac{\bar{Y}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$.

13. Ved test af $H_0 : \mu = \mu_0$ i modellen i opgave 11 benyttes testvariablen $\frac{\bar{Y}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$ under H_0.

    Opstil for alternativet $H_1 : \mu > \mu_0$ et udtryk for styrkefunktionen $\gamma(\mu)$.

    Facit: $\gamma(\mu) = \Phi\left(\frac{\mu-\mu_0}{\frac{\sigma}{\sqrt{n}}} - u_{1-\alpha}\right)$, hvor $u_{1-\alpha}$ er $(1-\alpha)$-fraktilen i den standardiserede normalfordeling.

    Vis, at et krav om at $\gamma(\mu) \geq \gamma$, hvor γ er en given værdi, fører til følgende betingelse om antallet af observationer:

$$n \geq \left(\frac{(u_{1-\alpha} - u_{1-\gamma})\sigma}{\mu - \mu_0}\right)^2 .$$

14. Bestem middelværdi og varians af $V \sim \chi^2(k)$. Bestem endvidere V's momentfrembringende funktion.

2 Simpel lineær regression

2.1 Modelbestemmelse

Lad $(x_1, y_1), \ldots, (x_n, y_n)$ være n sæt af observationer, dog således at x_i'erne er givne (valgte) konstanter, hvorimod y_i'erne er realiserede værdier af normalfordelte stokastiske variable med en middelværdi, der afhænger lineært af x_i, og med ens varians, dvs.

$$Y_i \sim N(\alpha + \beta x_i, \sigma^2) \ , \quad i = 1, \ldots, n \quad \text{uafhængige.}$$

Med $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$, hvor $\mu_i = \alpha + \beta x_i$, med $\mathbf{1} = (1, \ldots, 1)$ og med $\boldsymbol{x} = (x_1, \ldots, x_n)$ ser vi, at $\boldsymbol{\mu} \in L \subseteq R^n$, hvor $L = \operatorname{sp}\{\mathbf{1}, \boldsymbol{x}\}$. L er altså et todimensionalt underrum af R^n.

Modellen har tre parametre, α, β og σ^2, som skal estimeres.

En reparametrisering af modellen fås ved at sætte

$$\alpha + \beta x_i = \alpha + \beta\bar{x} + \beta(x_i - \bar{x}) = \alpha_0 + \beta t_i \ .$$

Her er det α_0, β og σ^2, der skal estimeres.

2.2 Mindste kvadraters estimater

Størrelsen, som skal minimeres, får her følgende udseende:

$$\|\boldsymbol{y} - \boldsymbol{\mu}\|^2 = \sum_i \big(y_i - (\alpha + \beta x_i)\big)^2 = \sum_i (y_i - \alpha - \beta x_i)^2$$

Formen er altså en funktion af de to reelle variable, α og β. Med benyttelse af den sædvanlige metode fra calculus til bestemmelse af lokale ekstrema får vi minimum for

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \quad \text{og} \quad \hat{\beta} = \frac{\sum_i x_i(y_i - \bar{y})}{\sum_i x_i(x_i - \bar{x})} \ ,$$

eller i den omformede model

$$\hat{\alpha_0} = \bar{y} \quad \text{og} \quad \hat{\beta} = \frac{\sum_i t_i y_i}{\sum_i t_i^2} \ .$$

Det er almindeligt at indføre forkortede betegnelser for bl.a. tæller og nævner i $\hat{\beta}$, men der er langt fra enighed om hvilke. For nævneren benyttes fx S_{xx}, SS_x, SSD_x, ssd_x, SAK_x m.fl. Her vil der blive benyttet følgende betegnelser:

$$\begin{aligned}
S_{xx} &= \sum_i (x_i - \bar{x})^2 \qquad \left(= \sum_i t_i^2 \right) \ , \\
S_{xy} &= \sum_i (x_i - \bar{x})(y_i - \bar{y}) \ , \\
S_{yy} &= \sum_i (y_i - \bar{y})^2 \ .
\end{aligned}$$

Bemærk, at $\sum_i x_i(y_i - \bar{y}) = \sum_i (x_i - \bar{x})y_i = \sum_i (x_i - \bar{x})(y_i - \bar{y})$. Denne og tilsvarende identiteter eftervises i opgave 2.

Med de forkortede betegnelser har vi

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \quad \text{og} \quad \hat{\beta} = \frac{S_{xy}}{S_{xx}} \ ,$$

subsidiært

$$\hat{\alpha_0} = \bar{y} \quad \text{og} \quad \hat{\beta} = \frac{S_{xy}}{S_{xx}} \ .$$

Både $\hat{\alpha}$ og $\hat{\beta}$ ses at være linearkombinationer af y_i'erne og er derfor som stokastiske variable normalfordelte, jf. afsnit 3.3.

For middelværdierne får vi:

$$\begin{aligned}
\mathrm{E}\left[\hat{\beta}\right] &= \frac{\mathrm{E}\left[S_{xY}\right]}{S_{xx}} = \frac{\sum_i (x_i - \bar{x})\mathrm{E}\left[Y_i\right]}{S_{xx}} = \frac{\sum_i (x_i - \bar{x})(\alpha + \beta x_i)}{S_{xx}} \\
&= \frac{\alpha \sum_i (x_i - \bar{x}) + \beta \sum_i (x_i - \bar{x}) x_i}{S_{xx}} = \frac{0 + \beta S_{xx}}{S_{xx}} = \beta \ , \\
\mathrm{E}\left[\hat{\alpha}\right] &= \mathrm{E}\left[\bar{Y}\right] - (\mathrm{E}\left[\hat{\beta}\right])\bar{x} = \frac{1}{n}\sum_i \mathrm{E}\left[Y_i\right] - \beta\bar{x} \\
&= \frac{1}{n}\sum_i (\alpha + \beta x_i) - \beta\bar{x} = \alpha + \beta\bar{x} - \beta\bar{x} = \alpha \ .
\end{aligned}$$

$\hat{\alpha}$ og $\hat{\beta}$ er altså centrale estimater for hhv. α og β.

Før vi bestemmer varianserne, vil vi vise, at $\bar{Y}$ og $\hat{\beta}$ er ukorrelerede og dermed uafhængige, da de begge er normalfordelte (jf. afsnit 3.3):

$$\begin{aligned}
\operatorname{Cov}\left(\bar{Y},\hat{\beta}\right) &= \operatorname{Cov}\left(\frac{1}{n}\sum_i Y_i, \frac{1}{S_{xx}}\sum_j (x_j-\bar{x})Y_j\right) \\
&= \frac{1}{nS_{xx}}\sum_i\sum_j (x_j-\bar{x})\operatorname{Cov}(Y_i,Y_j) \\
&= \frac{\sigma^2}{nS_{xx}}\sum_i (x_i-\bar{x}) = 0\ .
\end{aligned}$$

Varianserne bliver

$$\begin{aligned}
\operatorname{Var}\left(\hat{\beta}\right) &= \frac{\operatorname{Var}(S_{xY})}{{S_{xx}}^2} = \frac{\sum_i (x_i-\bar{x})^2\operatorname{Var}(Y_i)}{{S_{xx}}^2} = \frac{\sigma^2 S_{xx}}{{S_{xx}}^2} = \frac{\sigma^2}{S_{xx}}\ , \\
\operatorname{Var}(\hat{\alpha}) &= \operatorname{Var}\left(\bar{Y}\right) + (\operatorname{Var}\left(\hat{\beta}\right))\bar{x}^2 = \frac{\sigma^2}{n} + \frac{\sigma^2}{S_{xx}}\bar{x}^2 \\
&= \frac{\sigma^2(S_{xx}+n\bar{x}^2)}{nS_{xx}} = \frac{\sigma^2(\sum_i x_i^2 - n\bar{x}^2 + n\bar{x}^2)}{nS_{xx}} = \frac{\sigma^2\sum_i x_i^2}{nS_{xx}}\ .
\end{aligned}$$

Vi kan nu sammenfatte

$$\hat{\alpha} \sim N\left(\alpha\,, \frac{\sigma^2\sum_i x_i^2}{nS_{xx}}\right)\ , \qquad \hat{\beta} \sim N\left(\beta\,, \frac{\sigma^2}{S_{xx}}\right)\ ,$$

$$\operatorname{Cov}\left(\hat{\alpha},\hat{\beta}\right) = \operatorname{Cov}\left(\bar{Y}-\hat{\beta}\bar{x},\hat{\beta}\right) = 0 - \bar{x}\operatorname{Var}\left(\hat{\beta}\right) = -\frac{\sigma^2\bar{x}}{S_{xx}}\ .$$

I den omformede model får vi

$$\hat{\alpha_0} \sim N\left(\alpha_0\,, \frac{\sigma^2}{n}\right) \quad \text{og} \quad \hat{\beta} \sim N\left(\beta\,, \frac{\sigma^2}{S_{xx}}\right) \quad \text{uafhængige.}$$

2.3 Residualer og residualkvadratsum

Modellens residualer er

$$\begin{aligned}
r_i &= y_i - \hat{y}_i = y_i - (\hat{\alpha}+\hat{\beta}x_i) = y_i - (\bar{y} - \hat{\beta}\bar{x} + \hat{\beta}x_i) \\
&= y_i - \bar{y} - \hat{\beta}(x_i-\bar{x})\ , \quad i = 1,\ldots,n,
\end{aligned}$$

dvs. linearkombinationer af y_i'erne og dermed som stokastiske variable normalfordelte. Der gælder, at

$$\begin{aligned}
\operatorname{E}[R_i] &= \operatorname{E}[Y_i] - \operatorname{E}\left[\bar{Y}\right] - \operatorname{E}\left[\hat{\beta}\right](x_i-\bar{x}) \\
&= \alpha + \beta x_i - (\alpha + \beta\bar{x}) - \beta(x_i-\bar{x}) = 0\ .
\end{aligned}$$

Endvidere, at

$$\mathrm{Cov}\left(Y_i - \bar{Y}, \hat{\beta}(x_i - \bar{x})\right) = \frac{\sigma^2(x_i - \bar{x})^2}{S_{xx}} \ ,$$

jf. opgave 5.

$$\begin{aligned}
\mathrm{Var}\,(R_i) &= \mathrm{Var}\left(Y_i - \bar{Y}\right) - 2\mathrm{Cov}\left(Y_i - \bar{Y}, \hat{\beta}(x_i - \bar{x})\right) + \mathrm{Var}\left(\hat{\beta}(x_i - \bar{x})\right) \\
&= \sigma^2\left(1 - \frac{1}{n}\right) - 2\,\frac{\sigma^2(x_i - \bar{x})^2}{S_{xx}} + \frac{\sigma^2(x_i - \bar{x})^2}{S_{xx}} \\
&= \sigma^2\left(1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{S_{xx}}\right) \ .
\end{aligned}$$

De enkelte residualer er korrelerede, der gælder, at

$$\mathrm{Cov}\,(R_i, R_j) = -\sigma^2\left(\frac{1}{n} + \frac{(x_i - \bar{x})(x_j - \bar{x})}{S_{xx}}\right) \ , \quad i \neq j \ ,$$

jf. opgave 6.

Residualkvadratsummen kan udtrykkes som

$$\begin{aligned}
\sum_i r_i^2 &= \sum_i (y_i - \hat{y}_i)^2 = \sum_i (y_i - \hat{\alpha} - \hat{\beta}x_i)^2 \\
&= \sum_i (y_i - \bar{y} - \hat{\beta}(x_i - \bar{x}))^2 \\
&= \sum_i (y_i - \bar{y})^2 - 2\hat{\beta}\sum_i (x_i - \bar{x})(y_i - \bar{y}) + \hat{\beta}^2\sum_i (x_i - \bar{x})^2 \\
&= S_{yy} - 2\hat{\beta}S_{xy} + \hat{\beta}^2 S_{xx} \ .
\end{aligned}$$

Ved benyttelse af $\hat{\beta} = \frac{S_{xy}}{S_{xx}}$ kan vi opnå tre varianter af residualkvadratsummen:

$$\sum_i r_i^2 = S_{yy} - \hat{\beta}^2 S_{xx} = S_{yy} - \hat{\beta}S_{xy} = \frac{S_{xx}S_{yy} - {S_{xy}}^2}{S_{xx}} \ .$$

Med residualkvadratsummen som stokastisk variabel får vi

$$\mathrm{E}\left[\sum_i R_i^2\right] = \mathrm{E}\left[S_{YY}\right] - \mathrm{E}\left[\hat{\beta}^2\right] S_{xx} \ .$$

Efter en hel del regninger fremkommer

$$\mathrm{E}\left[\sum_i R_i^2\right] = (n-2)\sigma^2 \ ,$$

jf. opgave 7.

Vi har altså, at $\frac{1}{n-2}\sum_i r_i^2$ er et centralt estimat for σ^2. Denne størrelse betegnes normalt s^2.

Det kan vises[4], at

$$S^2 \sim \frac{\sigma^2}{n-2}\,\chi^2(n-2) \quad \text{uafhængig af } (\hat{\alpha}\,,\hat{\beta})\ .$$

S^2 er også uafhængig af $(\hat{\alpha_0},\hat{\beta})$, og da $\hat{\alpha_0}$ og $\hat{\beta}$ er uafhængige, gælder følgelig at $\hat{\alpha_0}$, $\hat{\beta}$ og S^2 er uafhængige. Heraf endvidere, at $\hat{\alpha}$ og S^2 er uafhængige.

Middelværdi og varians af S^2:

$$\begin{aligned}
\mathrm{E}\left[S^2\right] &= \frac{\sigma^2}{n-2}\,(n-2) = \sigma^2 \quad \text{(regnekontrol)},\\
\mathrm{Var}\left(S^2\right) &= \frac{\sigma^4}{(n-2)^2}\,2(n-2) = \frac{2\sigma^4}{n-2}\ .
\end{aligned}$$

2.4 Determinationskoefficienten r^2

Bemærk, at

$$\begin{aligned}
\sum_i (\hat{y}_i - \bar{y})^2 &= \sum_i (\hat{\alpha} + \hat{\beta}x_i - \bar{y})^2 = \sum_i (\bar{y} - \hat{\beta}\bar{x} + \hat{\beta}x_i - \bar{y})^2\\
&= \sum_i (\hat{\beta}(x_i - \bar{x}))^2 = \hat{\beta}^2 S_{xx} = \hat{\beta} S_{xy} = \frac{{S_{xy}}^2}{S_{xx}}\ ,
\end{aligned}$$

og betragt

$$r^2 = \frac{\sum_i (\hat{y}_i - \bar{y})^2}{S_{yy}} = \frac{{S_{xy}}^2}{S_{xx}S_{yy}}\ .$$

Vi kan tolke r^2 som den brøkdel af variationen i y-værdierne, som regressionsmodellen kan forklare. Betegnelsen r^2 er valgt, så r svarer til den empiriske korrelationskoefficient for x- og y-værdierne, idet der åbenbart gælder, at

$$r = \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}} = \frac{(n-1)s_{xy}}{\sqrt{(n-1){s_x}^2}\sqrt{(n-1){s_y}^2}} = \frac{s_{xy}}{s_x s_y}\ ,$$

hvor $s_x = \sqrt{\frac{1}{n-1}\sum_i (x_i - \bar{x})^2}$, $s_y = \sqrt{\frac{1}{n-1}\sum_i (y_i - \bar{y})^2}$ og
$s_{xy} = \frac{1}{n-1}\sum_i (x_i - \bar{x})(y_i - \bar{y})$ i overenstemmelse med sædvanlig notation.

Bemærk, at $r = \hat{\beta}\frac{s_x}{s_y}$, jf. opgave 8.

[4] Vi vender tilbage til spørgsmålet i opgave 11 i afsnit 4.12.

2.5 Maksimum likelihood estimater

Likelihoodfunktionen bliver

$$
\begin{aligned}
L(\alpha,\beta,\sigma^2) &= \prod_i \frac{1}{\sqrt{2\pi}\,\sigma}\ \mathrm{e}^{-\frac{(y_i-\alpha-\beta x_i)^2}{2\sigma^2}} \\
&= (2\pi\sigma^2)^{-\frac{n}{2}}\ \mathrm{e}^{-\frac{1}{2\sigma^2}\sum_i (y_i-\alpha-\beta x_i)^2}\ ,
\end{aligned}
$$

og dermed

$$
l(\alpha,\beta,\sigma^2) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_i (y_i-\alpha-\beta x_i)^2\ .
$$

For vilkårligt $\sigma^2 > 0$ antages den maksimale værdi af l, når $\sum_i (y_i-\alpha-\beta x_i)^2$ er mindst. Vi skal altså foretage præcis det samme valg af α og β, som MK-metoden anviste. Endvidere ser vi, at profilloglikelihooden for σ^2,

$$
\check{l}(\sigma^2) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_i (y_i-\hat{\alpha}-\hat{\beta} x_i)^2\ ,
$$

har samme form som den tilsvarende funktion i den simple lineære model, dvs. ML-estimatet bliver

$$
\hat{\sigma^2} = \frac{1}{n}\sum_i (y_i-\hat{\alpha}-\hat{\beta} x_i)^2
$$

og

$$
L(\hat{\alpha},\hat{\beta},\hat{\sigma^2}) = (2\pi e\hat{\sigma^2})^{-\frac{n}{2}}\ .
$$

Vi benytter normalt ikke $\hat{\sigma^2}$ men i stedet s^2 som defineret i forbindelse med MK-estimater. Sammenhængen mellem s^2 og $\hat{\sigma^2}$ er $s^2 = \frac{n}{n-2}\hat{\sigma^2}$.

2.6 Sufficient reduktion

Idet

$$
\sum_i (y_i-\alpha-\beta x_i)^2 = \sum_i ((y_i-\bar{y}) + \bar{y} - (\alpha+\beta x_i))^2
$$

$$
= S_{yy} + n\bar{y}^2 + n\alpha^2 + \beta^2\sum_i {x_i}^2 - 2\beta S_{xy} - 2n\alpha\bar{y} - 2n\beta\bar{x}\bar{y} + 2n\alpha\beta\bar{x}
$$

ses, at fx $(\bar{y}, S_{yy}, S_{xy})$ er sufficient for (α,β,σ^2).

2.7 Modelkontrol

For at opnå, at alle residualerne får samme varians, definerer vi de såkaldte standardiserede residualer

$$s_i = \frac{r_i}{\sqrt{1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{S_{xx}}}} \ , \quad i = 1, \ldots, n \ ,$$

hvorefter

$$S_i \sim N(0, \sigma^2) \ , \quad i = 1, \ldots, n \ .$$

Der ses ofte bort fra afhængigheden mellem de enkelte S_i'er. Kontrol for normalitet kan da udføres ved at optegne et normalfraktildiagram.

2.8 Konfidensintervaller

Ved at udnytte kendskabet til fordelingerne af $\hat{\alpha}$, $\hat{\beta}$ og S^2 kan vi umiddelbart opstille konfidensintervaller for parametrene α, β og σ^2:

$$\begin{aligned}
\alpha &= \hat{\alpha} \pm t_{1-\frac{\alpha}{2}}(n-2)\, s \sqrt{\frac{\sum_i x_i^2}{n S_{xx}}} \ , \\
\beta &= \hat{\beta} \pm t_{1-\frac{\alpha}{2}}(n-2)\, \frac{s}{\sqrt{S_{xx}}} \ , \\
\frac{(n-2)s^2}{\chi^2_{1-\frac{\alpha}{2}}(n-2)} \ &\leq \ \sigma^2 \ \leq \ \frac{(n-2)s^2}{\chi^2_{\frac{\alpha}{2}}(n-2)} \ ,
\end{aligned}$$

hvor der til alle tre intervaller er benyttet konfidensgraden $1 - \alpha$. Bemærk, at symbolet α traditionelt benyttes både som betegnelse for en parameter og i forbindelse med angivelse af konfidensgrad. Det burde ikke give anledning til misforståelser.

Betragter vi $\mu = \alpha + \beta x$, hvor x ikke nødvendigvis falder i en af x_i-værdierne, kan vi umiddelbart danne estimatet $\hat{\mu} = \hat{\alpha} + \hat{\beta} x$. $\hat{\mu}$ benævnes også $\hat{y}$.

$\hat{\mu}$ er et centralt estimat for μ, idet $\mathrm{E}[\hat{\mu}] = \mathrm{E}[\hat{\alpha}] + \mathrm{E}\left[\hat{\beta}\right] x = \alpha + \beta x = \mu$.

Bemærk, at $\hat{\mu} = \bar{y} - \hat{\beta}\bar{x} + \hat{\beta}x = \bar{y} + \hat{\beta}(x - \bar{x})$ og dermed, at $\mathrm{Var}(\hat{\mu}) = \frac{\sigma^2}{n} + \frac{\sigma^2}{S_{xx}}(x - \bar{x})^2$, idet $\bar{Y}$ og $\hat{\beta}$ er uafhængige.

Da $\hat{\mu}$ som linearkombination af normalfordelte variable selv er normalfordelt, har vi

$$\hat{\mu} \sim N\left(\alpha + \beta x \, , \sigma^2 \left(\frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}}\right)\right) \ .$$

Konfidensintervallet for μ med konfidensgrad $1 - \alpha$ bliver derfor

$$\mu = \hat{y} \pm t_{1-\frac{\alpha}{2}}(n-2)\, s \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}}} \ .$$

Vi kan opfatte konfidensgrænserne for μ som funktioner af x. Graferne kaldes for konfidensbåndet.

Vi vil også bestemme de såkaldte prædiktionsgrænser for en tænkt fremtidig observation y svarende til værdien x af den forklarende variabel. Idet $Y \sim N(\alpha+\beta x\,,\sigma^2)$ er uafhængig af tidligere observationer, må der gælde

$$Y-\hat{\mu} \sim N\left(0\,,\sigma^2\left(1+\frac{1}{n}+\frac{(x-\bar{x})^2}{S_{xx}}\right)\right) \ .$$

Benytter vi samme teknik som ved konstruktion af konfidensintervaller, får vi, at

$$P\left(-t_{1-\frac{\alpha}{2}}(n-2) \ \le \ \frac{Y-\hat{\mu}}{S\sqrt{1+\frac{1}{n}+\frac{(x-\bar{x})^2}{S_{xx}}}} \ \le \ t_{1-\frac{\alpha}{2}}(n-2)\right) = 1-\alpha \ ,$$

som omformes til

$$P\left(\hat{\mu}-t_{1-\frac{\alpha}{2}}(n-2)\,S\sqrt{1+\frac{1}{n}+\frac{(x-\bar{x})^2}{S_{xx}}} \ \le \ Y \ \le \right.$$
$$\left. \hat{\mu}+t_{1-\frac{\alpha}{2}}(n-2)\,S\sqrt{1+\frac{1}{n}+\frac{(x-\bar{x})^2}{S_{xx}}}\ \right) = 1-\alpha \ .$$

Heraf følger prædiktionsintervallet for y med konfidensgrad $1-\alpha$:

$$y = \hat{y} \pm t_{1-\frac{\alpha}{2}}(n-2)\,s\sqrt{1+\frac{1}{n}+\frac{(x-\bar{x})^2}{S_{xx}}} \ .$$

Også her kan vi bestemme et bånd ved at lade x variere. Bemærk, at med samme konfidensgrad er bredden her altid større end ved konfidensintervallet for μ.

2.9 Test i modellen

Kendskabet til parameterestimaternes fordelinger gør det muligt umiddelbart at opstille hypotesetest om parametrene.

Ønsker vi fx at teste $H_0 : \beta = \beta_0$, $H_1 : \beta \neq \beta_0$, kan vi benytte testvariablen

$$T = \frac{\hat{\beta}-\beta_0}{\frac{S}{\sqrt{S_{xx}}}} \sim t(n-2) \quad \text{under } H_0 \ ,$$

og dermed udføre et simpelt t-test. Bemærk, at

$$\frac{s}{\sqrt{S_{xx}}} = \sqrt{\frac{\frac{1}{n-2}\sum_i {r_i}^2}{S_{xx}}} = \frac{1}{S_{xx}}\sqrt{\frac{S_{xx}S_{yy}-{S_{xy}}^2}{n-2}}.$$

Ofte testes $H_0 : \beta = 0$ for at afklare om regressionsmodellen er relevant. I dette test kan den observerede værdi af testvariablen T udtrykkes alene ved determinationskoefficienten r^2:

$$t_{obs} = \frac{\hat{\beta}}{\frac{s}{\sqrt{S_{xx}}}} = \frac{\frac{S_{xy}}{S_{xx}}}{\frac{1}{S_{xx}}\sqrt{\frac{S_{xx}S_{yy}-{S_{xy}}^2}{n-2}}} = \frac{\frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}\sqrt{n-2}}{\sqrt{\frac{S_{xx}S_{yy}-{S_{xy}}^2}{S_{xx}S_{yy}}}} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \ .$$

Hypoteser om σ^2 testes ved χ^2-test, jf. opgave 11.

2.10 Opgaver

1. Gennemfør udregningerne til bestemmelse af $\hat{\alpha}$ og $\hat{\beta}$.

2. Eftervis følgende identiteter:

$$\sum_i (x_i - \bar{x})^2 = \sum_i (x_i - \bar{x})x_i = \sum_i {x_i}^2 - n\bar{x}^2 = \sum_i {x_i}^2 - \frac{(\sum_i x_i)^2}{n},$$

$$\begin{aligned}\sum_i (x_i - \bar{x})(y_i - \bar{y}) = \sum_i (x_i - \bar{x})y_i = \sum_i x_i(y_i - \bar{y}) &= \sum_i x_i y_i - n\bar{x}\bar{y} \\ &= \sum_i x_i y_i - \frac{\sum_i x_i \sum_i y_i}{n}.\end{aligned}$$

3. Eftervis, at $\hat{\beta}$ bestemt i den sædvanlige model og $\hat{\beta}$ bestemt i den reparametriserede model er identiske.

4. Vis, at $(\mathbf{1}, \boldsymbol{t})$, hvor $\boldsymbol{t} = \boldsymbol{x} - \bar{x}\mathbf{1}$, er en ortogonal basis for for L.

   Udnyt dette til at vise, at $\|\hat{\boldsymbol{\mu}}\|^2 = n\bar{y}^2 + \hat{\beta}^2 S_{xx}$.

5. Eftervis, at $\mathrm{Cov}\left(Y_i - \bar{Y}, \hat{\beta}(x_i - \bar{x})\right) = \frac{\sigma^2(x_i-\bar{x})^2}{S_{xx}}$.

6. Eftervis, at $\mathrm{Cov}\,(R_i, R_j) = -\sigma^2\left(\frac{1}{n} + \frac{(x_i-\bar{x})(x_j-\bar{x})}{S_{xx}}\right)$, $i \neq j$.

7. Eftervis, at $\mathrm{E}\left[\sum_i {R_i}^2\right] = (n-2)\sigma^2$. Vink: Benyt $\sum_i {R_i}^2$ på formen $S_{YY} - \hat{\beta}^2 S_{xx}$, og husk at $S_{YY} = \sum_i (Y_i - \bar{Y})^2$.

   1. metode: Sæt $Y_i = \alpha + \beta x_i + U_i$, $U_i \sim N(0, \sigma^2)$, $i = 1, \ldots, n$. Husk regnereglen $\mathrm{E}\,[X^2] = \mathrm{Var}\,(X) + (\mathrm{E}\,[X])^2$.

   2. metode: Udregn $(Y_i - \bar{Y})^2$. Benyt regnereglen for $\mathrm{E}\,[X^2]$ samt regnereglen $\mathrm{E}\,[XY] = \mathrm{Cov}\,(X,Y) + \mathrm{E}\,[X]\,\mathrm{E}\,[Y]$.

8. Eftervis, at $r = \hat{\beta}\frac{s_x}{s_y}$. Vink: Tag udgangspunkt i $r^2 = \frac{{S_{xy}}^2}{S_{xx}S_{yy}} = \hat{\beta}^2\frac{S_{xx}}{S_{yy}}$.

9. Eftervis, at $(\bar{y}, S_{yy}, S_{xy})$ er sufficient for $(\alpha, \beta, \sigma^2)$.

10. Angiv konfidensintervallet for parameteren $\bar{\mu} = \alpha + \beta\bar{x}$ med konfidensgrad $1 - \alpha$ (α benyttes her i to forskellige betydninger).

11. Opstil et signifikanstest for $H_0 : \sigma^2 = {\sigma_0}^2$, $H_1 : \sigma^2 > {\sigma_0}^2$, dvs. angiv testvariabel og dennes fordeling under H_0 samt det tilhørende kritiske område.

12. Betragt modellen

$$Y_i \sim N(\beta x_i, \sigma^2)\ , \quad i = 1, \ldots, n \text{ uafhængige.}$$

Estimer parametre, bestem deres fordelinger, opstil konfidensintervaller og test m.m.

3 Den flerdimensionale normalfordeling

3.1 Stokastiske vektorer

Ved en stokastisk vektor skal vi forstå en vektor, hvor de enkelte komponenter er sædvanlige stokastiske variable.

For den stokastiske vektor $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ definerer vi middelværdivektoren som

$$\mathrm{E}[\boldsymbol{Y}] = (\mathrm{E}[Y_1], \ldots, \mathrm{E}[Y_n])\ .$$

I sammenhæng med matricer vil vi opfatte $\boldsymbol{Y}$ og $\mathrm{E}[\boldsymbol{Y}]$ som søjlevektorer.

Lad os betragte to stokastiske vektorer $\boldsymbol{X} = (X_1, \ldots, X_m)$ og $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$. Alle kovarianser mellem $\boldsymbol{X}$'s komponenter og $\boldsymbol{Y}$'s komponenter kan naturligt samles i en $m \times n$ matrix, den såkaldte kovariansmatrix:

$$\begin{aligned}
\mathrm{Cov}(\boldsymbol{X}, \boldsymbol{Y}) &= \{\mathrm{Cov}(X_i, Y_j)\} \\
&= \{\mathrm{E}[(X_i - \mathrm{E}[X_i])(Y_j - \mathrm{E}[Y_j])]\} \\
&= \mathrm{E}\left[(\boldsymbol{X} - \mathrm{E}[\boldsymbol{X}])(\boldsymbol{Y} - \mathrm{E}[\boldsymbol{Y}])^\top\right]\ .
\end{aligned}$$

I det sidste udtryk skal vi ved middelværdien af en matrix naturligvis forstå matricen bestående af alle elementernes middelværdier.

Sætter vi $\boldsymbol{X} = \boldsymbol{Y}$ i $\mathrm{Cov}(\boldsymbol{X}, \boldsymbol{Y})$, får vi den såkaldte variansmatrix for $\boldsymbol{Y}$ (også kaldet varians-kovariansmatrix, kovariansmatrix, dispersionsmatrix):

$$\mathrm{Var}(\boldsymbol{Y}) = \mathrm{Cov}(\boldsymbol{Y}, \boldsymbol{Y}) = \{\mathrm{Cov}(Y_i, Y_j)\}\ .$$

Bemærk, at $\mathrm{Var}(\boldsymbol{Y})$ har $\mathrm{Var}(Y_i)$, $i = 1, \ldots, n$, i hoveddiagonalen. Bemærk endvidere, at $\mathrm{Var}(\boldsymbol{Y})$ er symmetrisk, da $\mathrm{Cov}(Y_i, Y_j) = \mathrm{Cov}(Y_j, Y_i)$.

Eksempel: Ved simpel lineær regression fandt vi om residualerne, at

$$R_i \sim N\left(0, \sigma^2\left(1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{S_{xx}}\right)\right),$$

og at

$$\mathrm{Cov}(R_i, R_j) = -\sigma^2\left(\frac{1}{n} + \frac{(x_i - \bar{x})(x_j - \bar{x})}{S_{xx}}\right), \quad i \neq j .$$

Sætter vi

$$p_{ij} = \frac{1}{n} + \frac{(x_i - \bar{x})(x_j - \bar{x})}{S_{xx}}$$

og

$$P = \{p_{ij}\} ,$$

gælder der åbenbart, at

$$\mathrm{Var}(\boldsymbol{R}) = \sigma^2(I - P) . \qquad \square$$

At to stokastiske vektorer er uafhængige, betyder at en vilkårlig komponent i den ene vektor er uafhængig af en vilkårlig komponent i den anden vektor. (To forskellige komponenter i samme vektor kan være afhængige eller uafhængige.) Bemærk, at

$$\boldsymbol{X}, \boldsymbol{Y} \text{ uafhængige} \quad \Rightarrow \quad \mathrm{Cov}(\boldsymbol{X}, \boldsymbol{Y}) = O.$$

Lad os betragte en lineær transformation T af den stokastiske vektor $\boldsymbol{Y}$ repræsenteret ved matricen A, dvs. $T : R^n \to R^m$, $T(\boldsymbol{Y}) = A\boldsymbol{Y}$. Da middelværdioperatoren er lineær, har vi umiddelbart, at $\mathrm{E}[A\boldsymbol{Y}] = A\,\mathrm{E}[\boldsymbol{Y}]$. For variansmatricen får vi, at

$$\begin{aligned}
\mathrm{Var}(A\boldsymbol{Y}) &= \mathrm{E}\left[(A\boldsymbol{Y} - A\mathrm{E}[\boldsymbol{Y}])(A\boldsymbol{Y} - A\mathrm{E}[\boldsymbol{Y}])^\top\right] \\
&= \mathrm{E}\left[A(\boldsymbol{Y} - \mathrm{E}[\boldsymbol{Y}])(\boldsymbol{Y} - \mathrm{E}[\boldsymbol{Y}])^\top A^\top\right] \\
&= A\,\mathrm{E}\left[(\boldsymbol{Y} - \mathrm{E}[\boldsymbol{Y}])(\boldsymbol{Y} - \mathrm{E}[\boldsymbol{Y}])^\top\right] A^\top \\
&= A\,\mathrm{Var}(\boldsymbol{Y})\,A^\top .
\end{aligned}$$

$\mathrm{Var}(A\boldsymbol{Y})$ er altså en $m \times m$ matrix.

Betragter vi specielt en linearkombination af $\boldsymbol{Y}$'s komponenter $a_1Y_1 + \ldots + a_nY_n$, der kort kan skrives $\boldsymbol{a}^\top\boldsymbol{Y}$, får vi $\mathrm{E}\left[\boldsymbol{a}^\top\boldsymbol{Y}\right] = \boldsymbol{a}^\top\mathrm{E}[\boldsymbol{Y}]$ og $\mathrm{Var}\left(\boldsymbol{a}^\top\boldsymbol{Y}\right) = \boldsymbol{a}^\top\mathrm{Var}(\boldsymbol{Y})\,\boldsymbol{a}$. $\mathrm{Var}\left(\boldsymbol{a}^\top\boldsymbol{Y}\right)$ er et reelt tal (1×1 matrix).

En ortogonal transformation er en lineær transformation, der repræsenteres ved en ortogonal matrix[5], dvs. en matrix hvor søjlevektorerne er ortonormerede. For en

[5] Den traditionelle betegnelse, en mere præcis betegnelse ville være ortonormal matrix.

sådan matrix gælder, at systemet af rækkevektorer ligeledes er ortonormeret, at den inverse matrix er lig den transponerede, og at den tilhørende determinant antager værdien 1 eller -1.

Lad $\boldsymbol{Z} = C\boldsymbol{Y}$ definere en ortogonal transformation. C er altså en ortogonal matrix. Der gælder, at

$$\sum_i Z_i^2 = \boldsymbol{Z}^\top \boldsymbol{Z} = (C\boldsymbol{Y})^\top C\boldsymbol{Y} = \boldsymbol{Y}^\top C^\top C\boldsymbol{Y} = \boldsymbol{Y}^\top \boldsymbol{Y} = \sum_i Y_i^2 \,,$$

dvs. kvadratsummen af en stokastisk vektors komponenter er invariant over for ortogonaltransformation.

Betragt nu en stokastisk kvadratisk form $\boldsymbol{Y}^\top A\boldsymbol{Y}$, hvor $\mathrm{E}\left[\boldsymbol{Y}\right] = \boldsymbol{\mu}$ og $\mathrm{Var}\left(\boldsymbol{Y}\right) = \Sigma$. Middelværdien af $\boldsymbol{Y}^\top A\boldsymbol{Y}$ udregnes:

$$\begin{aligned}
\mathrm{E}\left[\boldsymbol{Y}^\top A\boldsymbol{Y}\right] &= \mathrm{E}\left[\sum_i \sum_j a_{ij} Y_i Y_j\right] = \sum_i \sum_j a_{ij} \mathrm{E}\left[Y_i Y_j\right] \\
&= \sum_i \sum_j a_{ij}\left(\mathrm{Cov}\left(Y_i, Y_j\right) + \mathrm{E}\left[Y_i\right]\mathrm{E}\left[Y_j\right]\right) \\
&= \sum_i \sum_j a_{ij}\sigma_{ij} + \sum_i \sum_j a_{ij}\mu_i\mu_j \\
&= \sum_i \sum_j a_{ij}\sigma_{ji} + \boldsymbol{\mu}^\top A\boldsymbol{\mu} = \sum_i (A\Sigma)_{ii} + \boldsymbol{\mu}^\top A\boldsymbol{\mu} \\
&= \mathrm{tr}\, A\Sigma + \boldsymbol{\mu}^\top A\boldsymbol{\mu}
\end{aligned}$$

Når $\mathrm{Var}\left(\boldsymbol{Y}\right) = \sigma^2 I$ forenkles udtrykket til

$$\mathrm{E}\left[\boldsymbol{Y}^\top A\boldsymbol{Y}\right] = \sigma^2 \mathrm{tr}\, A + \boldsymbol{\mu}^\top A\boldsymbol{\mu} \,,$$

og når yderligere $\mathrm{E}\left[\boldsymbol{Y}\right] = \boldsymbol{0}$ til

$$\mathrm{E}\left[\boldsymbol{Y}^\top A\boldsymbol{Y}\right] = \sigma^2 \mathrm{tr}\, A \,.$$

Den momentfrembringende funktion hørende til en stokastisk vektor defineres som

$$M_{\boldsymbol{Y}}(t_1, \ldots, t_n) = \mathrm{E}\left[\exp(\boldsymbol{t}^\top \boldsymbol{Y})\right].$$

Ofte skrives $M_{\boldsymbol{Y}}(\boldsymbol{t})$ som en forkortelse for $M_{\boldsymbol{Y}}(t_1, \ldots, t_n)$.

De sædvanlige egenskaber ved momentfrembringende funktion har også gyldighed her, dvs.

$$M_{A\boldsymbol{Y}+\boldsymbol{b}}(\boldsymbol{t}) = \exp(\boldsymbol{t}^\top \boldsymbol{b}) M_{\boldsymbol{Y}}(A^\top \boldsymbol{t}),$$

og for $\boldsymbol{X}$ og $\boldsymbol{Y}$ uafhængige

$$M_{\boldsymbol{X}+\boldsymbol{Y}}(\boldsymbol{t}) = M_{\boldsymbol{X}}(\boldsymbol{t}) M_{\boldsymbol{Y}}(\boldsymbol{t}),$$

jf. opgave 4 og 5.

3.2 Den flerdimensionale normalfordeling

En stokastisk vektor $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ siges, at være n-dimensional normalfordelt, når $Y_1, \ldots, Y_n$ har den simultane tæthed

$$f_{\boldsymbol{Y}}(y_1, \ldots, y_n) = (2\pi)^{-\frac{n}{2}} (\det \Sigma)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} (\boldsymbol{y} - \boldsymbol{\mu})^\top \Sigma^{-1} (\boldsymbol{y} - \boldsymbol{\mu}) \right) ,$$

hvor $\boldsymbol{\mu} \in R^n$ og $\Sigma \in R^{n \times n}$. $\boldsymbol{\mu}$ kan vælges vilkårligt, mens Σ skal være en positiv definit matrix[6].

At $\boldsymbol{Y}$ er n-dimensional normalfordelt med parametrene $\boldsymbol{\mu}$ og Σ, skrives kort $\boldsymbol{Y} \sim N_n(\boldsymbol{\mu}, \Sigma)$.

Vi vil først vise, at vi ved en passende affin transformation af $\boldsymbol{Y}$ kan opnå, at komponenterne i den transformerede stokastiske vektor er uafhængige og standardiserede normalfordelte.

Sæt $\boldsymbol{Y} = \Sigma^{\frac{1}{2}} \boldsymbol{X} + \boldsymbol{\mu}$ (se opgave 7, 8 og 9 vedr. $\Sigma^{\frac{1}{2}}$). Jacobideterminanten hørende til denne transformation ses umiddelbart at være $\det \Sigma^{\frac{1}{2}}$. Tæthedsfunktionen for $\boldsymbol{X}$ bliver derfor

$$\begin{aligned} f_{\boldsymbol{X}}(x_1, \ldots, x_n) &= (2\pi)^{-\frac{n}{2}} (\det \Sigma)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \boldsymbol{x}^\top \Sigma^{\frac{1}{2}} \Sigma^{-1} \Sigma^{\frac{1}{2}} \boldsymbol{x} \right) \left| \det \Sigma^{\frac{1}{2}} \right| \\ &= (2\pi)^{-\frac{n}{2}} \exp\left(-\frac{1}{2} \boldsymbol{x}^\top \boldsymbol{x} \right) = (2\pi)^{-\frac{n}{2}} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} {x_i}^2 \right) \\ &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{{x_i}^2}{2}} , \end{aligned}$$

hvoraf ses, at $X_i \sim N(0, 1)$, $i = 1, \ldots, n$, og at X_i'erne er uafhængige.

Herefter kan vi let udregne $\mathrm{E}[\boldsymbol{Y}]$ og $\mathrm{Var}(\boldsymbol{Y})$:

$$\begin{aligned} \mathrm{E}[\boldsymbol{Y}] &= \Sigma^{\frac{1}{2}} \mathrm{E}[\boldsymbol{X}] + \boldsymbol{\mu} = \Sigma^{\frac{1}{2}} \boldsymbol{0} + \boldsymbol{\mu} = \boldsymbol{\mu} , \\ \mathrm{Var}(\boldsymbol{Y}) &= \Sigma^{\frac{1}{2}} \mathrm{Var}(\boldsymbol{X}) (\Sigma^{\frac{1}{2}})^\top = \Sigma^{\frac{1}{2}} I_n \Sigma^{\frac{1}{2}} = \Sigma . \end{aligned}$$

Parametrene $\boldsymbol{\mu}$ og Σ er altså henholdsvis middelværdivektor og variansmatrix for $\boldsymbol{Y}$.

Den momentfrembringende funktion for $\boldsymbol{Y}$ kan vi bestemme ved følgende udreg-

[6] En symmetrisk $n \times n$ matrix er positiv definit, når $\forall \boldsymbol{x} \in R^n \setminus \{\boldsymbol{0}\} : \boldsymbol{x}^\top A \boldsymbol{x} > 0$. A er positiv semidefinit, når $\forall \boldsymbol{x} \in R^n \setminus \{\boldsymbol{0}\} : \boldsymbol{x}^\top A \boldsymbol{x} \geq 0$, og der findes (mindst) et $\boldsymbol{x} \neq \boldsymbol{0}$, så $\boldsymbol{x}^\top A \boldsymbol{x} = 0$. Analogt defineres negativ definit og negativ semidefinit.

ninger, hvor vi igen benytter transformationen $\boldsymbol{Y} = \Sigma^{\frac{1}{2}}\boldsymbol{X} + \boldsymbol{\mu}$:

$$\begin{aligned}
M_{\boldsymbol{Y}}(\boldsymbol{t}) &= \mathrm{E}\left[\exp(\boldsymbol{t}^\top \boldsymbol{Y})\right] \\
&= \int_{R^n} \exp(\boldsymbol{t}^\top \boldsymbol{y})(2\pi)^{-\frac{n}{2}}(\det\Sigma)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\right) dV \\
&= \int_{R^n} \exp(\boldsymbol{t}^\top(\Sigma^{\frac{1}{2}}\boldsymbol{x}+\boldsymbol{\mu}))(2\pi)^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\boldsymbol{x}^\top\boldsymbol{x}\right) dV \\
&= \exp(\boldsymbol{t}^\top\boldsymbol{\mu}) \int_{R^n} (2\pi)^{-\frac{n}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}^\top\boldsymbol{x} - 2\boldsymbol{t}^\top\Sigma^{\frac{1}{2}}\boldsymbol{x})\right) dV \\
&= \exp\left(\boldsymbol{t}^\top\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}^\top\Sigma\boldsymbol{t}\right) \int_{R^n} (2\pi)^{-\frac{n}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}^\top\boldsymbol{x} - 2\boldsymbol{t}^\top\Sigma^{\frac{1}{2}}\boldsymbol{x} + \boldsymbol{t}^\top\Sigma\boldsymbol{t})\right) dV \\
&= \exp\left(\boldsymbol{t}^\top\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}^\top\Sigma\boldsymbol{t}\right) \int_{R^n} (2\pi)^{-\frac{n}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \Sigma^{\frac{1}{2}}\boldsymbol{t})^\top(\boldsymbol{x} - \Sigma^{\frac{1}{2}}\boldsymbol{t})\right) dV \\
&= \exp\left(\boldsymbol{t}^\top\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}^\top\Sigma\boldsymbol{t}\right) ,
\end{aligned}$$

idet funktionen under det sidste integraltegn ses at være tæthedsfunktionen for en n-dimensional normalfordeling med middelværdivektor $\Sigma^{\frac{1}{2}}\boldsymbol{t}$ og variansmatrix I_n.

I stedet for at definere den flerdimensionale normalfordeling ud fra tæthedsfunktionen kan vi benytte følgende alternative definition:

$\boldsymbol{Y}$ er n-dimensional normalfordelt, når og kun når der for alle $\boldsymbol{l} \neq \boldsymbol{0}$ gælder, at $\boldsymbol{l}^\top\boldsymbol{Y}$ er (endimensional) normalfordelt. Benytter vi symbolerne $\boldsymbol{\mu}$ og Σ som ovenfor, skal der altså gælde at $\boldsymbol{l}^\top\boldsymbol{Y} \sim N(\boldsymbol{l}^\top\boldsymbol{\mu}, \boldsymbol{l}^\top\Sigma\,\boldsymbol{l})$.

I denne definition kan vi umiddelbart medtage det singulære tilfælde, hvor Σ kun er positiv semidefinit svarende til, at der findes (mindst) et $\boldsymbol{l} \neq \boldsymbol{0}$, så $\boldsymbol{l}^\top\Sigma\,\boldsymbol{l} = 0$. I det singulære tilfælde vil der altså optræde linearkombination(er) $\boldsymbol{l}^\top\boldsymbol{Y}$ med varians 0.

Fra tidligere kender vi den momentfrembringende funktion for en normalfordelt stokastisk variabel $X \sim N(\mu, \sigma^2)$. Der gælder $M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$. Følgelig har vi

$$M_{\boldsymbol{l}^\top\boldsymbol{Y}}(t) = \exp\left(\boldsymbol{l}^\top\boldsymbol{\mu}t + \frac{1}{2}\boldsymbol{l}^\top\Sigma\,\boldsymbol{l}t^2\right) .$$

Ved benyttelse af momentfrembringende funktioner kan det let eftervises, at de to forskellige definitioner på flerdimensional normalfordeling er ækvivalente, når der ses bort fra det singulære tilfælde, se opgave 11.

3.3 Egenskaber ved normalfordelte stokastiske vektorer

Lad os betragte en affin transformation $A\boldsymbol{Y}+\boldsymbol{b}$ af $\boldsymbol{Y} \sim N_n(\boldsymbol{\mu}\,, \Sigma)$, hvor A er $q \times n$ og rang $A = q \leq n$. Der gælder

$$
\begin{aligned}
\forall \boldsymbol{l} \neq \boldsymbol{0} \;:\; \boldsymbol{l}^\top(A\boldsymbol{Y}+\boldsymbol{b}) &= (\boldsymbol{l}^\top A)\boldsymbol{Y} + \boldsymbol{l}^\top \boldsymbol{b} = \boldsymbol{l}_1^\top \boldsymbol{Y} + b_1 \\
&\sim N(\boldsymbol{l}_1^\top \boldsymbol{\mu} + b_1, \boldsymbol{l}_1^\top \Sigma\, \boldsymbol{l}_1) \\
&= N(\boldsymbol{l}^\top A\boldsymbol{\mu} + \boldsymbol{l}^\top \boldsymbol{b}, \boldsymbol{l}^\top A\Sigma A^\top \boldsymbol{l}) \;,
\end{aligned}
$$

som viser, at $A\boldsymbol{Y}+\boldsymbol{b} \sim N_q(A\boldsymbol{\mu}+\boldsymbol{b}, A\Sigma A^\top)$, idet $A\Sigma A^\top$ er positiv definit, jf. opgave 12.

Heraf følger, at en vilkårlig linearkombination $\boldsymbol{a}^\top\boldsymbol{Y}$ af $\boldsymbol{Y}$'s komponenter er normalfordelt.

Ved et passende valg af A kan vi udtage en vilkårlig delvektor af $\boldsymbol{Y}$. Alle delvektorer af $\boldsymbol{Y}$, herunder også de enkelte komponenter, er altså normalfordelte.

For endimensionale normalfordelte variable gælder, at uafhængighed er ensbetydende med, at de variable er ukorrelerede. Betragter vi to normalfordelte stokastiske vektorer $\boldsymbol{Y}_1 \sim N_{n_1}(\boldsymbol{\mu}_1\,, \Sigma_1)$ og $\boldsymbol{Y}_2 \sim N_{n_2}(\boldsymbol{\mu}_2\,, \Sigma_2)$, vil vi tilsvarende vise, at uafhængighed mellem $\boldsymbol{Y}_1$ og $\boldsymbol{Y}_2$ er ensbetydende med, at $\mathrm{Cov}\,(\boldsymbol{Y}_1, \boldsymbol{Y}_2) = O$.

Det gælder generelt, at uafhængighed mellem stokastiske vektorer medfører, at de er ukorrelerede, jf. afsnit 3.1.

Antager vi omvendt, at $\mathrm{Cov}\,(\boldsymbol{Y}_1, \boldsymbol{Y}_2) = O$, bliver variansmatricen Σ for den stokastiske vektor $\boldsymbol{Y} = (\boldsymbol{Y}_1, \boldsymbol{Y}_2)$

$$
\Sigma = \begin{bmatrix} \Sigma_1 & O \\ O & \Sigma_2 \end{bmatrix},
$$

hvor Σ_1 og Σ_2 er variansmatricerne for henholdsvis $\boldsymbol{Y}_1$ og $\boldsymbol{Y}_2$. Heraf følger, at

$$
\Sigma^{-1} = \begin{bmatrix} {\Sigma_1}^{-1} & O \\ O & {\Sigma_2}^{-1} \end{bmatrix}.
$$

Bemærk, at

$$
\begin{aligned}
(\boldsymbol{y}-\boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{y}-\boldsymbol{\mu}) &= \begin{bmatrix} \boldsymbol{y}_1 - \boldsymbol{\mu}_1 \\ \boldsymbol{y}_2 - \boldsymbol{\mu}_2 \end{bmatrix}^\top \begin{bmatrix} {\Sigma_1}^{-1} & O \\ O & {\Sigma_2}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{y}_1 - \boldsymbol{\mu}_1 \\ \boldsymbol{y}_2 - \boldsymbol{\mu}_2 \end{bmatrix} \\
&= (\boldsymbol{y}_1 - \boldsymbol{\mu}_1)^\top \Sigma_1^{-1}(\boldsymbol{y}_1 - \boldsymbol{\mu}_1) + (\boldsymbol{y}_2 - \boldsymbol{\mu}_2)^\top \Sigma_2^{-1}(\boldsymbol{y}_2 - \boldsymbol{\mu}_2) \\
&= h_1(\boldsymbol{y}_1) + h_2(\boldsymbol{y}_2)
\end{aligned}
$$

For $\boldsymbol{Y}$'s tæthedsfunktion får vi derfor, at

$$
\begin{aligned}
&f_{\boldsymbol{Y}}(y_{11},\dots,y_{1n_1},y_{21},\dots,y_{2n_2})\\
&\quad=(2\pi)^{-\frac{n_1+n_2}{2}}(\det\Sigma_1\det\Sigma_2)^{-\frac{1}{2}}\exp\big(-\tfrac{1}{2}(h_1(\boldsymbol{y}_1)+h_2(\boldsymbol{y}_2))\big)\\
&\quad=(2\pi)^{-\frac{n_1}{2}}(\det\Sigma_1)^{-\frac{1}{2}}\exp\big(-\tfrac{1}{2}h_1(\boldsymbol{y}_1)\big)(2\pi)^{-\frac{n_2}{2}}(\det\Sigma_2)^{-\frac{1}{2}}\exp\big(-\tfrac{1}{2}h_2(\boldsymbol{y}_2)\big)\\
&\quad=f_{\boldsymbol{Y}_1}(y_{11},\dots,y_{1n_1})f_{\boldsymbol{Y}_2}(y_{21},\dots,y_{2n_2}),
\end{aligned}
$$

hvoraf det ses, at $\boldsymbol{Y}_1$ og $\boldsymbol{Y}_2$ er uafhængige.

Hvis vi betragter to lineære transformationer af samme vektor $\boldsymbol{Y}$, $A_1\boldsymbol{Y}$ og $A_2\boldsymbol{Y}$, hvor A_1 og A_2 har fuld rang, ses umiddelbart, at uafhængighed mellem $A_1\boldsymbol{Y}$ og $A_2\boldsymbol{Y}$ er ensbetydende med at $A_1\Sigma A_2^\top=O$, idet

$$\mathrm{Cov}\,(A_1\boldsymbol{Y},A_2\boldsymbol{Y})=O\Leftrightarrow A_1\mathrm{Cov}\,(\boldsymbol{Y},\boldsymbol{Y})\,A_2^\top=O\Leftrightarrow A_1\Sigma A_2^\top=O.$$

Bemærk, at med $A_1\Sigma A_2^\top=O$ er systemet af rækker i A_1 ortogonalt på systemet af rækker i A_2 med hensyn til det indre produkt $\langle\boldsymbol{u},\boldsymbol{v}\rangle=\boldsymbol{u}^\top\Sigma\boldsymbol{v}$. Matricen $A=\begin{bmatrix}A_1\\A_2\end{bmatrix}$ har derfor fuld rang, dvs. $\operatorname{rang}A=\operatorname{rang}A_1+\operatorname{rang}A_2\le n$.

Specielt gælder, at to linearkombinationer af $\boldsymbol{Y}$'s komponenter, $\boldsymbol{a}_1{}^\top\boldsymbol{Y}$ og $\boldsymbol{a}_2{}^\top\boldsymbol{Y}$, er uafhængige, når og kun når $\mathrm{Cov}\,\big(\boldsymbol{a}_1{}^\top\boldsymbol{Y},\boldsymbol{a}_2{}^\top\boldsymbol{Y}\big)=0\Leftrightarrow\boldsymbol{a}_1{}^\top\Sigma\boldsymbol{a}_2=0$.

Et vigtigt resultat er, at

$$(\boldsymbol{Y}-\boldsymbol{\mu})^\top\Sigma^{-1}(\boldsymbol{Y}-\boldsymbol{\mu})\sim\chi^2(n)\ ,$$

hvilket let eftervises, se opgave 14.

Lad os til slut vende tilbage til påstanden i afsnit 1.4, om at $\bar{Y}$ og S^2 er uafhængige. $\bar{Y}$ og S^2 er baseret på en normalfordelt stikprøve med middelværdi μ og varians σ^2, dvs. $\boldsymbol{Y}\sim N_n(\mu\mathbf{1},\sigma^2 I)$.

Når $\mathbf{Z}$ er en ortogonal transformation af $\boldsymbol{Y}$, dvs. $\boldsymbol{Z}=C\boldsymbol{Y}$ med C ortogonal, får vi

$$\boldsymbol{Z}\sim N_n(\mu C\mathbf{1},\sigma^2 CIC^\top)=N_n(\mu C\mathbf{1},\sigma^2 I)\,.$$

Komponenterne i $\boldsymbol{Z}$ er altså uafhængige.

Benytter vi specielt en ortogonal matrix, der har vektoren $(\frac{1}{\sqrt{n}},\dots,\frac{1}{\sqrt{n}})$ som første rækkevektor, får vi

$$Z_1=\begin{bmatrix}\frac{1}{\sqrt{n}}&\cdots&\frac{1}{\sqrt{n}}\end{bmatrix}\boldsymbol{Y}=\frac{1}{\sqrt{n}}\sum_i Y_i=\sqrt{n}\ \bar{Y}\ \ \Leftrightarrow\ \ \bar{Y}=\frac{Z_1}{\sqrt{n}}$$

og

$$
\begin{aligned}
(n-1)S^2&=\sum_i(Y_i-\bar{Y})^2=\sum_i Y_i^2-n\bar{Y}^2=\sum_i Z_i^2-n\left(\frac{Z_1}{\sqrt{n}}\right)^2\\
&=Z_2^2+\cdots+Z_n^2\,,
\end{aligned}
$$

hvoraf det umiddelbart fremgår, at $\bar{Y}$ og S^2 er uafhængige.

3.4 En betinget fordeling

Lad $\boldsymbol{Y} \sim N_n(\boldsymbol{\mu}, \Sigma)$, og betragt delvektorene $\boldsymbol{Y}_1$ med n_1 komponenter og $\boldsymbol{Y}_2$ med n_2 komponenter, $n_1 + n_2 = n$. $\boldsymbol{Y}_1$ og $\boldsymbol{Y}_2$ er begge normalfordelte og i almindelighed afhængige. Vi kan skrive

$$(\boldsymbol{Y}_1, \boldsymbol{Y}_2) \sim N_{n_1+n_2}\left((\boldsymbol{\mu}_1, \boldsymbol{\mu}_2), \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right),$$

hvor betydningen af de benyttede betegnelser er åbenbar, fx står Σ_{12} for $\operatorname{Cov}(\boldsymbol{Y}_1, \boldsymbol{Y}_2)$.

Indfør $\boldsymbol{Z} = \boldsymbol{Y}_2 - \Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{Y}_1$, og betragt den stokastiske vektor $(\boldsymbol{Y}_1, \boldsymbol{Z})$, der fremkommer som en lineær transformation af $(\boldsymbol{Y_1}, \boldsymbol{Y_2})$. Transformationen er repræsenteret ved matricen $\begin{bmatrix} I_{n_1} & O \\ -\Sigma_{21}\Sigma_{11}^{-1} & I_{n_2} \end{bmatrix}$, som er $n \times n$ og regulær.

Ved udregning, jf. opgave 16, får vi, at

$$\operatorname{Var}((\boldsymbol{Y}_1, \boldsymbol{Z})) = \begin{bmatrix} \Sigma_{11} & O \\ O & \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \end{bmatrix},$$

der viser, at $\boldsymbol{Y}_1$ og $\boldsymbol{Z}$ er uafhængige.

For $\operatorname{Var}(\boldsymbol{Z})$ benyttes som regel betegnelsen $\Sigma_{22\cdot 1}$, altså

$$\Sigma_{22\cdot 1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}.$$

Bemærk, at

$$\mathrm{E}[\boldsymbol{Z}] = \mathrm{E}[\boldsymbol{Y}_2] - \Sigma_{21}\Sigma_{11}^{-1}\mathrm{E}[\boldsymbol{Y}_1] = \boldsymbol{\mu}_2 - \Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{\mu}_1,$$

hvorefter vi kan angive $\boldsymbol{Z}$'s fordeling som

$$\boldsymbol{Z} \sim N_{n_2}\left(\boldsymbol{\mu}_2 - \Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{\mu}_1, \Sigma_{22\cdot 1}\right).$$

Fra indførelsen af $\boldsymbol{Z}$ ser vi, at

$$\boldsymbol{Y}_2 = \boldsymbol{Z} + \Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{Y}_1,$$

og ved at betinge med $\boldsymbol{Y}_1 = \boldsymbol{y}_1$, får vi

$$\boldsymbol{Y}_2|\boldsymbol{y}_1 = \boldsymbol{Z} + \Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{y}_1,$$

idet $\boldsymbol{Y}_1$ og $\boldsymbol{Z}$ er uafhængige. Middelværdivektor og variansmatrix for $\boldsymbol{Y}_2|\boldsymbol{y}_1$ bliver

$$\mathrm{E}[\boldsymbol{Y}_2|\boldsymbol{y}_1] = \boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\boldsymbol{y}_1 - \boldsymbol{\mu}_1),$$

$$\operatorname{Var}(\boldsymbol{Y}_2|\boldsymbol{y}_1) = \operatorname{Var}(\boldsymbol{Z}) = \Sigma_{22\cdot 1}.$$

Fordelingen af $\boldsymbol{Y}_2|\boldsymbol{y}_1$ er altså

$$\boldsymbol{Y}_2|\boldsymbol{y}_1 \sim N_{n_2}\left(\boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\boldsymbol{y}_1 - \boldsymbol{\mu}_1), \Sigma_{22\cdot 1}\right).$$

3.5 Opgaver

1. Vis, at $\mathrm{Cov}(\boldsymbol{X}, \boldsymbol{Y}) = \mathrm{E}\left[\boldsymbol{X}\boldsymbol{Y}^\top\right] - \mathrm{E}[\boldsymbol{X}]\,\mathrm{E}[\boldsymbol{Y}]^\top$.

2. Vis, at $\mathrm{Cov}(A\boldsymbol{X}, B\boldsymbol{Y}) = A\,\mathrm{Cov}(\boldsymbol{X}, \boldsymbol{Y})\,B^\top$.

3. Vis, at $\mathrm{Var}(\boldsymbol{Y} + \boldsymbol{b}) = \mathrm{Var}(\boldsymbol{Y})$.

4. Vis, at $M_{A\boldsymbol{Y}+\boldsymbol{b}}(\boldsymbol{t}) = \exp(\boldsymbol{t}^\top\boldsymbol{b})M_{\boldsymbol{Y}}(A^\top\boldsymbol{t})$.

5. Vis, at $\boldsymbol{X}$ og $\boldsymbol{Y}$ uafhængige $\Rightarrow \quad M_{\boldsymbol{X}+\boldsymbol{Y}}(\boldsymbol{t}) = M_{\boldsymbol{X}}(\boldsymbol{t})M_{\boldsymbol{Y}}(\boldsymbol{t})$.

6. Lad C være en ortogonal matrix, dvs. der gælder $C^\top C = I$. Vis, at systemet af rækkevektorer i C er ortonormeret, at $C^{-1} = C^\top$, og at $\det C = \pm 1$. Vis endvidere, at produktet af to ortogonale matricer giver en ortogonal matrix.

7. Lad A være en symmetrisk matrix. Vis, at A er positiv definit, når og kun når alle A's egenværdier er positive. Vink: Betragt en ortogonal substitution $\boldsymbol{x} = C\boldsymbol{y}$, hvor matricen C diagonaliserer A, dvs. $C^{-1}AC = \Lambda = \ulcorner\lambda_1 \ldots \lambda_n\lrcorner$, og vis, at $\boldsymbol{x}^\top A\boldsymbol{x} = \sum_i \lambda_i y_i^2$.

8. Lad A være en positiv definit matrix, og lad C være en ortogonal matrix, som diagonaliserer A til Λ. Bestem først en matrix $\Lambda^{\frac{1}{2}}$, så $(\Lambda^{\frac{1}{2}})^2 = \Lambda$, og dernæst en matrix $A^{\frac{1}{2}}$, så $(A^{\frac{1}{2}})^2 = A$.'

9. Vis, at $\det \Sigma^{\frac{1}{2}} = (\det \Sigma)^{\frac{1}{2}}$.

10. Kontroller at $f_{\boldsymbol{Y}}$ er en tæthedsfunktion, dvs. at $\int_{R^n} f_{\boldsymbol{Y}}(y_1, \ldots, y_n)dV = 1$. Vink: Benyt tæthedsfunktionen $f_{\boldsymbol{X}}(x_1, \ldots, x_n)$ for den transformerede stokastiske vektor $\boldsymbol{X}$.

11. Vis ved benyttelse af momentfrembringende funktioner, at de to definitioner på flerdimensional normalfordeling er ækvivalente, dvs. at definition 1 $\Rightarrow$ definition 2, og at definition 2 $\Rightarrow$ definition 1. (Der ses bort fra det singulære tilfælde.)

12. Vis, at når A er $n \times n$ positiv definit, og når C er $p \times n$ med $\operatorname{rang} C = p$, så er $CAC^\top$ positiv definit.

13. Eftervis, at $\boldsymbol{u}^\top\Sigma\boldsymbol{v}$, hvor Σ er positiv definit, definerer et indre produkt.

14. Vis, at $(\boldsymbol{Y} - \boldsymbol{\mu})^\top\Sigma^{-1}(\boldsymbol{Y} - \boldsymbol{\mu}) \sim \chi^2(n)$. Vink: Benyt transformationen $\boldsymbol{Y} = \Sigma^{\frac{1}{2}}\boldsymbol{X} + \boldsymbol{\mu}$.

15. Lad $Y_1, \ldots, Y_n$ være en stikprøve i en normalfordelt population. Giv et nyt bevis for, at $S^2 \sim \frac{\sigma^2}{n-1}\chi^2(n-1)$. Vink: Betragt først $\boldsymbol{X} = \frac{1}{\sigma}(\boldsymbol{Y} - \boldsymbol{\mu})$ og dernæst $\boldsymbol{U} = C\boldsymbol{X}$ med samme C som i slutningen af afsnit 3.3.

16. Eftervis, at $\boldsymbol{Y_1}$ og $\boldsymbol{Z}$ er uafhængige. Vink: Husk, at Var$(A\boldsymbol{Y})$=AVar$(\boldsymbol{Y})A^\top$.

4 En generel lineær normal model

4.1 Modelbestemmelse

Lad som sædvanlig $y_1, \dots, y_n$ være observationer i en normalfordelt population, hvor alle variable har samme varians.

Vi har $(y_1, \dots, y_n) \in R^n$ og

$$Y_i \sim N(\mu_i, \sigma^2)\,, \quad i = 1, \dots, n \quad \text{uafhængige.}$$

Antag om middelværdivektoren $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)$, at $\boldsymbol{\mu} \in L \subseteq R^n$, hvor L er et underrum af R^n. Lad os endvidere antage, at L har dimension k og udspændes af vektorerne $\boldsymbol{x_1}, \dots, \boldsymbol{x_k}$, dvs.

$$L = \mathrm{sp}\,\{\boldsymbol{x_1}, \dots, \boldsymbol{x_k}\}\,, \quad \dim L = k\ .$$

Da $\boldsymbol{\mu} \in L$ findes konstanter $\beta_1, \dots, \beta_k$, så $\boldsymbol{\mu} = \beta_1 \boldsymbol{x_1} + \ldots + \beta_k \boldsymbol{x_k}$, eller

$$\boldsymbol{\mu} = X\boldsymbol{\beta}\,, \quad X = [\boldsymbol{x_1}\ \dots\ \boldsymbol{x_k}]\,, \quad \boldsymbol{\beta} = (\beta_1, \dots, \beta_k)\ .$$

I udtrykket $\boldsymbol{\mu} = X\boldsymbol{\beta}$ skal $\boldsymbol{\mu}$ og $\boldsymbol{\beta}$ opfattes som søjlevektorer. X kaldes modellens designmatrix.

Modellen har $k+1$ parametre, $\beta_1, \dots, \beta_k$ og σ^2, der alle skal estimeres.

4.2 Mindste kvadraters estimater

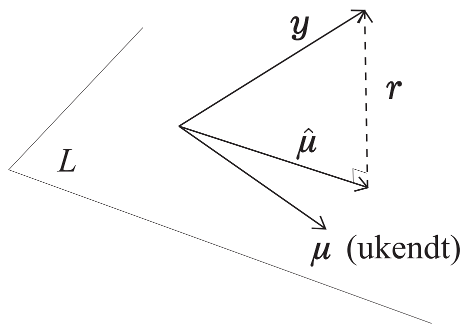
Lad $\hat{\boldsymbol{\mu}}$ betegne den ortogonale projektion af $\boldsymbol{y}$ på L, dvs. $\hat{\boldsymbol{\mu}} = P\boldsymbol{y}$, hvor P er projektionsmatrix.

Benyttelse af mindste kvadraters metode giver

$$\begin{aligned} \|\boldsymbol{y} - \boldsymbol{\mu}\|^2 &= \|\boldsymbol{y} - \hat{\boldsymbol{\mu}} + \hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|^2 \\ &= \|\boldsymbol{y} - \hat{\boldsymbol{\mu}}\|^2 + \|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|^2 \quad \text{(Pythagoras)} \\ &\geq \|\boldsymbol{y} - \hat{\boldsymbol{\mu}}\|^2\,, \end{aligned}$$

hvor lighedstegnet gælder for $\boldsymbol{\mu} = \hat{\boldsymbol{\mu}}$.

$\hat{\boldsymbol{\mu}} = P\boldsymbol{y}$ er altså mindste kvadraters estimat for $\boldsymbol{\mu}$.



Figur 3: Ortogonalprojektion af $\boldsymbol{y}$ på L.

$\hat{\boldsymbol{\mu}}$ er et centralt estimat for $\boldsymbol{\mu}$, idet

$$\mathrm{E}\,[\hat{\boldsymbol{\mu}}] = \mathrm{E}\,[P\boldsymbol{Y}] = P\mathrm{E}\,[\boldsymbol{Y}] = P\boldsymbol{\mu} = \boldsymbol{\mu}\ .$$

Søger vi et estimat $\hat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}$, skal der gælde, at $X\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\mu}}$, hvoraf følger at

$$X^\top X\hat{\boldsymbol{\beta}} = X^\top\hat{\boldsymbol{\mu}} = X^\top P\boldsymbol{y} = (P^\top X)^\top\boldsymbol{y} = (PX)^\top\boldsymbol{y} = X^\top\boldsymbol{y}\ .$$

Her har vi udnyttet, at en projektionsmatrix er symmetrisk. (Den anden grundlæggende egenskab ved en projektionsmatrix er, at den er idempotent.)

Vi ser, at $\hat{\boldsymbol{\beta}}$ skal opfylde ligningssystemet

$$X^\top X\hat{\boldsymbol{\beta}} = X^\top\boldsymbol{y}\ ,$$

som også kaldes normalligningerne.

Omvendt, når normalligningerne er opfyldt, gælder at

$$\begin{aligned} X^\top X\hat{\boldsymbol{\beta}} = X^\top\boldsymbol{y} \;\;&\Leftrightarrow\;\; X^\top\big(\boldsymbol{y} - X\hat{\boldsymbol{\beta}}\big) = \mathbf{0} \quad\Rightarrow\quad \hat{\boldsymbol{\beta}}^\top X^\top\big(\boldsymbol{y} - X\hat{\boldsymbol{\beta}}\big) = 0 \\ &\Leftrightarrow\;\; \big(X\hat{\boldsymbol{\beta}}\big)^\top\big(\boldsymbol{y} - X\hat{\boldsymbol{\beta}}\big) = 0 \quad\Leftrightarrow\quad X\hat{\boldsymbol{\beta}} \perp \boldsymbol{y} - X\hat{\boldsymbol{\beta}}\ , \end{aligned}$$

idet $\boldsymbol{y} = X\hat{\boldsymbol{\beta}} + \big(\boldsymbol{y} - X\hat{\boldsymbol{\beta}}\big)$ åbenbart er en ortogonal dekomposition af $\boldsymbol{y}$, hvor $X\hat{\boldsymbol{\beta}} \in L$ og $\boldsymbol{y} - X\hat{\boldsymbol{\beta}} \in L^\perp$. Da en ortogonal dekomposition er entydig, har vi, at $X\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\mu}}$ $(= P\boldsymbol{y})$.

Når $\operatorname{rang} X = k$, bliver $X^\top X$ en regulær $k \times k$ matrix, og normalligningerne kan umiddelbart løses:

$$\hat{\boldsymbol{\beta}} = (X^\top X)^{-1}X^\top\boldsymbol{y}\ .$$

Multiplicerer vi med X på begge sider får vi

$$X\hat{\boldsymbol{\beta}} = X(X^\top X)^{-1}X^\top\boldsymbol{y} \quad\Leftrightarrow\quad \hat{\boldsymbol{\mu}} = X(X^\top X)^{-1}X^\top\boldsymbol{y}.$$

Sammenholdes med $\hat{\boldsymbol{\mu}} = P\boldsymbol{y}$, ses at projektionsmatricen P kan udtrykkes som

$$P = X(X^\top X)^{-1}X^\top,$$

idet en projektionsmatrix er entydigt bestemt.

Hvis $\boldsymbol{x_1}, \ldots, \boldsymbol{x_k}$ er lineært afhængige, dvs. $\dim L = \operatorname{rang} X < k$, så vil $\hat{\boldsymbol{\mu}}$ og P stadig være entydigt bestemt, men ikke $\hat{\boldsymbol{\beta}}$.

Udtrykkene for $\hat{\boldsymbol{\beta}}$ og P kan opretholdes, hvis $(X^\top X)^{-1}$ erstattes af en vilkårlig generaliseret invers[7] $(X^\top X)^-$.

I det følgende forudsættes dog, at $\operatorname{rang} X = k$.

Med benyttelse af den flerdimensionale normalfordeling, kan vi formulere modellen som

$$\boldsymbol{Y} \sim N_n(\boldsymbol{\mu}, \sigma^2 I)\,, \quad \boldsymbol{\mu} = X\boldsymbol{\beta}\,, \quad X \quad n \times k\,, \quad \operatorname{rang} X = k\ .$$

Da $\hat{\boldsymbol{\beta}}$ er en lineær transformation af $\boldsymbol{y}$, er $\hat{\boldsymbol{\beta}} \sim N_k(\cdot,\cdot)$. Middelværdivektoren bliver

$$\begin{aligned} \mathrm{E}\,[\hat{\boldsymbol{\beta}}] &= (X^\top X)^{-1}X^\top \mathrm{E}\,[\boldsymbol{Y}] = (X^\top X)^{-1}X^\top \boldsymbol{\mu} \\ &= (X^\top X)^{-1}X^\top X\boldsymbol{\beta} = \boldsymbol{\beta}\ , \end{aligned}$$

dvs. $\hat{\boldsymbol{\beta}}$ er et centralt estimat for $\boldsymbol{\beta}$.

For variansmatricen gælder, at

$$\begin{aligned} \mathrm{Var}\,(\hat{\boldsymbol{\beta}}) &= (X^\top X)^{-1}X^\top \mathrm{Var}\,(\boldsymbol{Y})((X^\top X)^{-1}X^\top)^\top \\ &= (X^\top X)^{-1}X^\top \sigma^2 I X((X^\top X)^{-1})^\top \\ &= \sigma^2(X^\top X)^{-1}X^\top X(X^\top X)^{-1} \\ &= \sigma^2(X^\top X)^{-1}\ . \end{aligned}$$

Vi har altså

$$\hat{\boldsymbol{\beta}} \sim N_k(\boldsymbol{\beta}, \sigma^2(X^\top X)^{-1})\ .$$

4.3 Residualer og residualkvadratsum

Residualvektoren i modellen er

$$\begin{aligned} \boldsymbol{r} &= \boldsymbol{y} - \hat{\boldsymbol{y}} = \boldsymbol{y} - \hat{\boldsymbol{\mu}} = \boldsymbol{y} - X\hat{\boldsymbol{\beta}} \\ &= \boldsymbol{y} - P\boldsymbol{y} = (I-P)\boldsymbol{y}\ . \end{aligned}$$

Middelværdivektor og variansmatrix af $\boldsymbol{R}$ bliver

$$\begin{aligned} \mathrm{E}\,[\boldsymbol{R}] &= \mathrm{E}\,[\boldsymbol{Y}] - P\mathrm{E}\,[\boldsymbol{Y}] = \boldsymbol{\mu} - P\boldsymbol{\mu} = \boldsymbol{\mu} - \boldsymbol{\mu} = \boldsymbol{0}\ , \\ \mathrm{Var}\,(\boldsymbol{R}) &= (I-P)\mathrm{Var}\,(\boldsymbol{Y})(I-P)^\top = (I-P)\sigma^2 I(I-P) \end{aligned}$$

[7] A^- er en generaliseret invers til A, når der gælder $AA^-A = A$.

$$= \sigma^2(I-P)\ .$$

Tilsyneladende gælder $\boldsymbol{R} \sim N_n(\cdot,\cdot)$, men $I-P$ er positiv semidefinit, idet fx $\boldsymbol{\mu}^\top(I-P)\boldsymbol{\mu} = \boldsymbol{\mu}^\top(\boldsymbol{\mu}-\boldsymbol{\mu}) = \boldsymbol{\mu}^\top\mathbf{0} = 0$, dvs.

$$\boldsymbol{R} \sim N_n(\mathbf{0}, \sigma^2(I-P)) \quad \text{singulær.}$$

Da både $\hat{\boldsymbol{\beta}}$ og $\boldsymbol{R}$ er lineære transformationer af $\boldsymbol{Y}$, kan vi let udregne deres kovariansmatrix:

$$\begin{aligned}
\mathrm{Cov}\,(\hat{\boldsymbol{\beta}},\, \boldsymbol{R}) &= (X^\top X)^{-1}X^\top \mathrm{Var}\,(\boldsymbol{Y})(I-P)^\top \\
&= (X^\top X)^{-1}X^\top \sigma^2 I(I-P) \\
&= \sigma^2((X^\top X)^{-1}X^\top - (X^\top X)^{-1}X^\top P) \\
&= \sigma^2 O \\
&= O\ ,
\end{aligned}$$

idet $X^\top P = (P^\top X)^\top = (PX)^\top = X^\top$. Der gælder altså, at $\hat{\boldsymbol{\beta}}$ og $\boldsymbol{R}$ er uafhængige.

Residualkvadratsummen er

$$\|\boldsymbol{r}\|^2 = \|\boldsymbol{y}-\hat{\boldsymbol{\mu}}\|^2 = \|\boldsymbol{y}\|^2 - \|\hat{\boldsymbol{\mu}}\|^2 = \left(\boldsymbol{y} - X\hat{\boldsymbol{\beta}}\right)^\top \boldsymbol{y}\ ,$$

jf. opgave 6 og 7.

Residualkvadratsummen kan også skrives

$$((I-P)\boldsymbol{y})^\top(I-P)\boldsymbol{y} = \boldsymbol{y}^\top(I-P)\boldsymbol{y}\ .$$

Middelværdien kan med benyttelse af regneregel udledt i afsnit 3.1 udregnes til

$$\begin{aligned}
\mathrm{E}\left[\|\boldsymbol{R}\|^2\right] &= \mathrm{E}\left[\boldsymbol{Y}^\top(I-P)\boldsymbol{Y}\right] \\
&= \mathrm{tr}\,((I-P)\sigma^2 I) + \boldsymbol{\mu}^\top(I-P)\boldsymbol{\mu} \\
&= \sigma^2\mathrm{tr}\,(I-P) + \boldsymbol{\mu}^\top(\boldsymbol{\mu}-\boldsymbol{\mu}) \\
&= \sigma^2(\mathrm{tr}\,I - \mathrm{tr}\,P) + 0 \\
&= \sigma^2(n-k)\ ,
\end{aligned}$$

idet $\mathrm{tr}\,P = k$, jf. opgave 8.

Det ses, at $\frac{1}{n-k}\|\boldsymbol{r}\|^2 = \frac{1}{n-k}\left(\boldsymbol{y} - X\hat{\boldsymbol{\beta}}\right)^\top\boldsymbol{y}$ er et centralt estimat for σ^2. Vi vil benytte betegnelsen s^2 for dette estimat.

Ved at sætte $\boldsymbol{u} = \boldsymbol{y} - \boldsymbol{\mu}$, dvs. $\boldsymbol{U} \sim N_n(\mathbf{0}, \sigma^2 I)$ og indsætte $\boldsymbol{y} = \boldsymbol{u} + \boldsymbol{\mu}$ i udtrykket for residualkvadratsummen får vi endnu et udtryk for denne:

$$\boldsymbol{y}^\top(I-P)\boldsymbol{y} = (\boldsymbol{u}+\boldsymbol{\mu})^\top(I-P)(\boldsymbol{u}+\boldsymbol{\mu}) = \boldsymbol{u}^\top(I-P)\boldsymbol{u}\ ,$$

idet $(I-P)\boldsymbol{\mu} = \boldsymbol{\mu} - \boldsymbol{\mu} = \mathbf{0}$.

Lemma: Lad $\boldsymbol{U} \sim N_n(\mathbf{0}, \sigma^2 I)$, og lad P være en $n \times n$ projektionsmatrix med rang k. Der gælder da, at

$$\boldsymbol{U}^\top P \boldsymbol{U} \sim \sigma^2 \chi^2(k) \ .$$

Bevis: Da P er en projektionsmatrix, findes en ortogonal matrix C, så

$$C^{-1}PC = \Lambda = \lceil \underbrace{1 \ldots 1}_{k} \underbrace{0 \ldots 0}_{n-k} \rfloor \ ,$$

jf. opgave 9.

Betragt følgende lineære transformation af $\boldsymbol{U}$:

$$\begin{aligned}
[I_k \ O]\, C^{-1} P \boldsymbol{U} \quad &\sim \quad N_k(\mathbf{0}, [I_k \ O]\, C^{-1} P \mathrm{Var}\,(\boldsymbol{U})\, P (C^{-1})^\top \begin{bmatrix} I_k \\ O \end{bmatrix}) \\
&= \quad N_k(\mathbf{0}, \sigma^2\, [I_k \ O]\, C^{-1} P C \begin{bmatrix} I_k \\ O \end{bmatrix}) \\
&= \quad N_k(\mathbf{0}, \sigma^2\, [I_k \ O]\, \Lambda \begin{bmatrix} I_k \\ O \end{bmatrix}) \\
&= \quad N_k(\mathbf{0}, \sigma^2 I_k)
\end{aligned}$$

I regningerne blev benyttet, at $C^{-1} = C^\top$, og at $PIP = P^2 = P$.

Vi har nu umiddelbart, at

$$\begin{aligned}
&\left([I_k \ O]\, C^{-1} P \boldsymbol{U}\right)^\top \left(\sigma^2 I_k\right)^{-1} [I_k \ O]\, C^{-1} P \boldsymbol{U} \sim \chi^2(k) \\
\Leftrightarrow \ &\left([I_k \ O]\, C^{-1} P \boldsymbol{U}\right)^\top [I_k \ O]\, C^{-1} P \boldsymbol{U} \sim \sigma^2 \chi^2(k) \ ,
\end{aligned}$$

men

$$\begin{aligned}
&\left([I_k \ O]\, C^{-1} P \boldsymbol{U}\right)^\top [I_k \ O]\, C^{-1} P \boldsymbol{U} \\
&= \boldsymbol{U}^\top P C \begin{bmatrix} I_k \\ O \end{bmatrix} [I_k \ O]\, C^{-1} P \boldsymbol{U} \\
&= \boldsymbol{U}^\top P C \Lambda C^{-1} P \boldsymbol{U} \\
&= \boldsymbol{U}^\top P P P \boldsymbol{U} \\
&= \boldsymbol{U}^\top P \boldsymbol{U} \ ,
\end{aligned}$$

hvoraf følger, at $\boldsymbol{U}^\top P \boldsymbol{U} \sim \sigma^2 \chi^2(k)$. □

Ved benyttelse af ovenstående lemma får vi, at residualkvadratsummen

$$\|\boldsymbol{R}\|^2 = (n-k)S^2 = \boldsymbol{Y}^\top (I-P) \boldsymbol{Y} = \boldsymbol{U}^\top (I-P) \boldsymbol{U} \sim \sigma^2 \chi^2(n-k) \ ,$$

og dermed at

$$S^2 \sim \frac{\sigma^2}{n-k}\chi^2(n-k) \ .$$

Middelværdi og varians af S^2 bliver

$$\begin{aligned} \mathrm{E}\left[S^2\right] &= \frac{\sigma^2}{n-k}(n-k) = \sigma^2 \qquad \text{(regnekontrol)}, \\ \mathrm{Var}\left(S^2\right) &= \frac{\sigma^4}{(n-k)^2}\, 2(n-k) = \frac{2\sigma^4}{n-k} \ . \end{aligned}$$

Idet S^2 er en funktion af $\boldsymbol{R}$, har vi umiddelbart, at $\hat{\boldsymbol{\beta}}$ og S^2 er uafhængige.

4.4 Determinationskoefficienten R^2

Den brøkdel af variationen i observationerne, som modellen kan forklare, kaldes også i denne model for determinationskoefficienten, men betegnes her traditionelt med R^2, dvs.

$$R^2 = \frac{S_{yy} - \|\boldsymbol{r}\|^2}{S_{yy}} = \frac{\boldsymbol{y}^\top \boldsymbol{y} - n\bar{y}^2 - \left(\boldsymbol{y} - X\hat{\boldsymbol{\beta}}\right)^\top \boldsymbol{y}}{S_{yy}} = \frac{\left(X\hat{\boldsymbol{\beta}} - \bar{y}\mathbf{1}\right)^\top \boldsymbol{y}}{S_{yy}} \ .$$

S_{yy} har samme betydning som i afsnit 2.2.

4.5 Gauss-Markov

Blandt alle lineære centrale estimater af $\boldsymbol{c}^\top \boldsymbol{\mu}$ er $\boldsymbol{c}^\top \hat{\boldsymbol{\mu}}$ det estimat, der har mindst varians. Dette resultat kaldes Gauss-Markovs sætning.

En anden formulering er, at $\boldsymbol{c}^\top \hat{\boldsymbol{\mu}}$ er BLUE for $\boldsymbol{c}^\top \boldsymbol{\mu}$, hvor BLUE = Best Linear Unbiased Estimate. Med bedst menes naturligvis med mindst varians.

Det ses umiddelbart, at $\boldsymbol{c}^\top \hat{\boldsymbol{\mu}}$ er et lineært centralt estimat for $\boldsymbol{c}^\top \boldsymbol{\mu}$, idet $\boldsymbol{c}^\top \hat{\boldsymbol{\mu}} = \boldsymbol{c}^\top P\boldsymbol{Y} = (P\boldsymbol{c})^\top \boldsymbol{Y}$ og $\mathrm{E}\left[\boldsymbol{c}^\top \hat{\boldsymbol{\mu}}\right] = \boldsymbol{c}^\top \mathrm{E}\left[\hat{\boldsymbol{\mu}}\right] = \boldsymbol{c}^\top \boldsymbol{\mu}$.

Lad her $\boldsymbol{d}^\top \boldsymbol{Y}$ være et vilkårligt lineært centralt estimat for $\boldsymbol{c}^\top \boldsymbol{\mu}$.

Først udtrykker vi $\boldsymbol{c}^\top \hat{\boldsymbol{\mu}}$ ved $\boldsymbol{d}$ og $\boldsymbol{Y}$:

$$\begin{aligned} &\boldsymbol{c}^\top \boldsymbol{\mu} = \mathrm{E}\left[\boldsymbol{d}^\top \boldsymbol{Y}\right] = \boldsymbol{d}^\top \mathrm{E}\left[\boldsymbol{Y}\right] = \boldsymbol{d}^\top \boldsymbol{\mu} \\ &\Rightarrow (\boldsymbol{c} - \boldsymbol{d})^\top \boldsymbol{\mu} = 0 \ \Rightarrow \ \boldsymbol{c} - \boldsymbol{d} \in L^\perp \ \Rightarrow \ P(\boldsymbol{c} - \boldsymbol{d}) = \mathbf{0} \\ &\Rightarrow P\boldsymbol{c} = P\boldsymbol{d} \ \Rightarrow \ (P\boldsymbol{c})^\top \boldsymbol{Y} = (P\boldsymbol{d})^\top \boldsymbol{Y} \ \Rightarrow \ \boldsymbol{c}^\top \hat{\boldsymbol{\mu}} = (P\boldsymbol{d})^\top \boldsymbol{Y} \ . \end{aligned}$$

Herefter kan vi sammenholde varianserne af $\boldsymbol{c}^\top \hat{\boldsymbol{\mu}}$ og $\boldsymbol{d}^\top \boldsymbol{Y}$:

$$\mathrm{Var}\left(\boldsymbol{c}^\top \hat{\boldsymbol{\mu}}\right) = (P\boldsymbol{d})^\top \sigma^2 I P\boldsymbol{d} = \sigma^2 \boldsymbol{d}^\top P \boldsymbol{d} \ .$$

$$\mathrm{Var}\left(\boldsymbol{d}^\top \boldsymbol{Y}\right) = \boldsymbol{d}^\top \sigma^2 I \boldsymbol{d} = \sigma^2 \boldsymbol{d}^\top \boldsymbol{d} \ .$$

$$\mathrm{Var}\left(\boldsymbol{d}^\top \boldsymbol{Y}\right) - \mathrm{Var}\left(\boldsymbol{c}^\top \hat{\boldsymbol{\mu}}\right) = \sigma^2 \boldsymbol{d}^\top (I-P)\boldsymbol{d} \geq 0 \ , \quad \text{da } I-P \text{ er positiv semidefinit}$$

$$\Rightarrow \mathrm{Var}\left(\boldsymbol{d}^\top \boldsymbol{Y}\right) \geq \mathrm{Var}\left(\boldsymbol{c}^\top \hat{\boldsymbol{\mu}}\right) \ .$$

Entydighed ses af

$$\mathrm{Var}\left(\boldsymbol{d}^\top \boldsymbol{Y}\right) - \mathrm{Var}\left(\boldsymbol{c}^\top \hat{\boldsymbol{\mu}}\right) = 0 \ \Leftrightarrow \ \boldsymbol{d}^\top (I-P)\boldsymbol{d} = 0$$

$$\Leftrightarrow (I-P)\boldsymbol{d} = 0 \ \Leftrightarrow \ \boldsymbol{d} = P\boldsymbol{d} \ \Leftrightarrow \ \boldsymbol{d} = P\boldsymbol{c} \ .$$

Som korollar får vi – med et passende valg af $\boldsymbol{c}$ – at $\hat{\mu}_i$ er BLUE for μ_i.

4.6 Maksimum likelihood estimater

Likelihoodfunktionen bliver

$$\begin{aligned} L(\beta_1, \ldots, \beta_k, \sigma^2) &= \prod_i \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(y_i - \mu_i)^2}{2\sigma^2}\right) \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\|\boldsymbol{y} - \boldsymbol{\mu}\|^2\right) \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\|\boldsymbol{y} - X\boldsymbol{\beta}\|^2\right), \end{aligned}$$

og loglikelihoodfunktionen

$$l(\beta_1, \ldots, \beta_k, \sigma^2) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\|\boldsymbol{y} - X\boldsymbol{\beta}\|^2 \ .$$

For vilkårligt $\sigma^2 > 0$ har l maksimum når $\|\boldsymbol{y} - X\boldsymbol{\beta}\|^2 = \|\boldsymbol{y} - \boldsymbol{\mu}\|^2$ har minimum, dvs. netop for mindste kvadraters estimatet $\hat{\boldsymbol{\mu}} = X\hat{\boldsymbol{\beta}}$.

Profilloglikelihooden for σ^2 bliver

$$\check{l}(\sigma^2) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\|\boldsymbol{y} - X\hat{\boldsymbol{\beta}}\|^2 \ .$$

Denne funktion kan opfattes som en funktion af den reelle variabel σ^2. Maksimum bestemmes efter sædvanlig metode, og resultatet bliver

$$\hat{\sigma^2} = \frac{1}{n}\|\boldsymbol{y} - X\hat{\boldsymbol{\beta}}\|^2 \ .$$

Med dette udtryk for $\hat{\sigma^2}$ bliver likelihoodfunktionens maksimale værdi

$$L(\hat{\beta}_1, \ldots, \hat{\beta}_k, \hat{\sigma^2}) = (2\pi e \hat{\sigma^2})^{-\frac{n}{2}} \ .$$

Som i de tidligere modeller vil vi normalt ikke benytte $\hat{\sigma^2}$, men i stedet det tilhørende centrale estimat

$$s^2 = \frac{1}{n-k}\|\boldsymbol{r}\|^2 = \frac{1}{n-k}\|\boldsymbol{y} - \hat{\boldsymbol{\mu}}\|^2 = \frac{1}{n-k}\|\boldsymbol{y} - X\hat{\boldsymbol{\beta}}\|^2 = \frac{1}{n-k}\big(\boldsymbol{y} - X\hat{\boldsymbol{\beta}}\big)^\top \boldsymbol{y} \ .$$

Relationen mellem s^2 og $\hat{\sigma^2}$ er $s^2 = \frac{n}{n-k}\hat{\sigma^2}$.

4.7 Sufficient reduktion

Idet

$$\begin{aligned} \|\boldsymbol{y} - \boldsymbol{\mu}\|^2 &= \|\boldsymbol{y} - X\boldsymbol{\beta}\|^2 = (\boldsymbol{y} - X\boldsymbol{\beta})^\top(\boldsymbol{y} - X\boldsymbol{\beta}) \\ &= \boldsymbol{y}^\top\boldsymbol{y} - 2\boldsymbol{y}^\top X\boldsymbol{\beta} + \boldsymbol{\beta}^\top X^\top X\boldsymbol{\beta} \ , \end{aligned}$$

ser vi, at $(\boldsymbol{y}^\top\boldsymbol{y}, \boldsymbol{y}^\top X)$ er sufficient for $(\boldsymbol{\beta}, \sigma^2)$.

4.8 Modelkontrol

For de enkelte residualer gælder, at

$$R_i \sim N(0, \sigma^2(1 - p_{ii})) \ , \quad i = 1, \ldots, n \ ,$$

og

$$\mathrm{Cov}\,(R_i, R_j) = -\sigma^2 p_{ij} \ , \quad i \neq j \ ,$$

hvor p_{ij} betyder det ij'te element i projektionsmatricen $P = X(X^\top X)^{-1}X^\top$.

Vi indfører de standardiserede residualer

$$s_i = \frac{r_i}{\sqrt{1 - p_{ii}}} \ , \quad i = 1, \ldots, n \ ,$$

hvorefter

$$S_i \sim N(0, \sigma^2) \ , \quad i = 1, \ldots, n \ ,$$

og

$$\mathrm{Cov}\,(S_i, S_j) = -\sigma^2 \frac{p_{ij}}{\sqrt{1 - p_{ii}}\sqrt{1 - p_{jj}}} \ , \quad i \neq j \ .$$

Idet der ofte ses bort fra afhængigheden mellem S_i'erne, kan kontrol for normalitet udføres i form af et normalfraktildiagram.

4.9 Konfidensellipsoider og konfidensintervaller

Med $\hat{\boldsymbol{\beta}} \sim N_k(\boldsymbol{\beta},\, \sigma^2(X^\top X)^{-1})$ har vi, at

$$(X^\top X)^{\frac{1}{2}}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sim N_k(\mathbf{0},\, \sigma^2 I) \ ,$$

og

$$(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top X^\top X(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sim \sigma^2\chi^2(k) \ ,$$

jf. opgave 13.

Da $S^2 \sim \frac{\sigma^2}{n-k}\chi^2(n-k)$ er uafhængig af $\boldsymbol{\beta}$, kan vi umiddelbart konstruere en F-fordelt variabel

$$F = \frac{\frac{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})^\top X^\top X(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})}{\sigma^2 k}}{\frac{\frac{(n-k)S^2}{\sigma^2}}{n-k}} = \frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top X^\top X(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{k\, S^2} \sim F(k,\, n-k) \ .$$

Området

$$\left\{\boldsymbol{\beta} \in R^k \,\middle|\, \frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top X^\top X(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{k\, s^2} \leq f_{1-\alpha}(k,\, n-k)\right\}$$

kaldes konfidensellipsoiden for $\boldsymbol{\beta}$ med konfidensgrad $1-\alpha$.

Betragter vi de enkelte parametre, kan vi med udgangspunkt i

$$\hat{\beta}_i \sim N(\beta_i,\, \sigma^2 {(X^\top X)^{-1}}_{ii}) \ , \quad i = 1, \ldots, n \ ,$$

opskrive konfidensintervaller med konfidensgrad $1-\alpha$ for de enkelte β_i'ere:

$$\beta_i = \hat{\beta}_i \pm t_{1-\frac{\alpha}{2}}(n-k)s\sqrt{{(X^\top X)^{-1}}_{ii}} \ .$$

Konfidensintervallet for σ^2 bestemmes til

$$\frac{(n-k)s^2}{\chi^2_{1-\frac{\alpha}{2}}(n-k)} \leq \ \sigma^2 \ \leq \frac{(n-k)s^2}{\chi^2_{\frac{\alpha}{2}}(n-k)} \ ,$$

ligeledes med konfidensgrad $1-\alpha$.

Lad os betragte linearkombinationen $\mu = \boldsymbol{x}^\top\boldsymbol{\beta}$, hvor $\boldsymbol{x} \in R^k$. Denne afledte parameter estimeres ved $\hat{\mu} = \boldsymbol{x}^\top\hat{\boldsymbol{\beta}}$, og der gælder

$$\hat{\mu} \sim N\big(\mu,\, \sigma^2\boldsymbol{x}^\top(X^\top X)^{-1}\boldsymbol{x}\big) \ .$$

Vi kan derfor opstille et konfidensinterval for μ med konfidensgrad $1-\alpha$:

$$\mu = \hat{\mu} \pm t_{1-\frac{\alpha}{2}}(n-k)s\sqrt{\boldsymbol{x}^\top(X^\top X)^{-1}\boldsymbol{x}} \ ,$$

hvor $\hat{\mu} = \boldsymbol{x}^\top\hat{\boldsymbol{\beta}}$ og $s^2 = \frac{1}{n-k}(\boldsymbol{y} - X\hat{\boldsymbol{\beta}})^\top\boldsymbol{y}$, $S^2 \sim \frac{\sigma^2}{n-k}\chi^2(n-k)$.

Da $\hat{\mu}$ kan opfattes som en i middel prædikteret værdi for observationer foretaget ved $\boldsymbol{x}$, betegnes størrelsen ofte med $\hat{y}$. Vi kan derfor også skrive

$$\mu = \hat{y} \pm t_{1-\frac{\alpha}{2}}(n-k)s\sqrt{\boldsymbol{x}^\top (X^\top X)^{-1}\boldsymbol{x}} \ .$$

Vi vil ligeledes opstille et prædiktionsinterval for en tænkt fremtidig observation y foretaget ved $\boldsymbol{x}$. Der må gælde, at $Y \sim N(\boldsymbol{x}^\top \boldsymbol{\beta}, \sigma^2)$, og at Y og $\hat{\mu}$ er uafhængige, da den fremtidige observation ikke afhænger af de tidligere foretagne. Heraf

$$\begin{aligned} Y - \hat{\mu} &\sim N\big(\boldsymbol{x}^\top \boldsymbol{\beta} - \boldsymbol{x}^\top \boldsymbol{\beta}, \sigma^2 + \sigma^2 \boldsymbol{x}^\top (X^\top X)^{-1}\boldsymbol{x}\big) \\ &= N\big(0, \sigma^2(1 + \boldsymbol{x}^\top (X^\top X)^{-1}\boldsymbol{x})\big) \ . \end{aligned}$$

Med benyttelse af samme regneteknik som i afsnit 2.8 får vi prædiktionsintervallet for y med konfidensgrad $1-\alpha$ til

$$y = \hat{y} \pm t_{1-\frac{\alpha}{2}}(n-k)s\sqrt{1 + \boldsymbol{x}^\top (X^\top X)^{-1}\boldsymbol{x}} \ ,$$

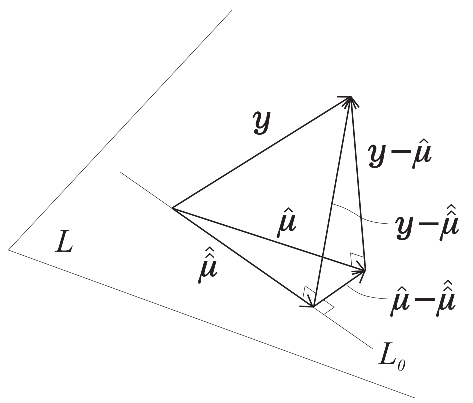
hvor $\hat{y} = \boldsymbol{x}^\top \hat{\boldsymbol{\beta}}$ og s som ovenfor.

Bemærk, at med samme konfidensgrad vil prædiktionsintervallet altid være større end det tilsvarende konfidensinterval.

4.10 Kvotienttest

Lad L_0 være et underrum af L, og betragt hypotesen $H_0 : \boldsymbol{\mu} \in L_0$ mod alternativet $H_1 : \boldsymbol{\mu} \in L \backslash L_0$.

Estimatet for $\boldsymbol{\mu}$ under H_0 skal naturligvis findes som projektionen af $\boldsymbol{y}$ på L_0. Vi betegner dette estimat $\hat{\hat{\boldsymbol{\mu}}}$. Figur 4 illustrerer sammenhængen mellem $\boldsymbol{y}$, $\hat{\boldsymbol{\mu}}$ og $\hat{\hat{\boldsymbol{\mu}}}$.



Figur 4: Sammenhængen mellem $\boldsymbol{y}$, $\hat{\boldsymbol{\mu}}$ og $\hat{\hat{\boldsymbol{\mu}}}$

Med benyttelse af resultater fra afsnit 4.6 kan likelihoodfunktionens maksimale værdi under H_0 angives til $\left(2\pi e\hat{\hat{\sigma}}^2\right)^{-\frac{n}{2}}$, hvor $\hat{\hat{\sigma}}^2 = \frac{1}{n}\|\boldsymbol{y} - \hat{\hat{\boldsymbol{\mu}}}\|^2$ og $\hat{\hat{\boldsymbol{\mu}}} = P_0\boldsymbol{y}$. P_0 er projektionsmatricen for den ortogonale projektion på L_0, $\operatorname{rang} P_0 = k_0 < k$.

Kvotientteststørrelsen $q(y_1, \ldots, y_n)$ bliver

$$\begin{aligned} q(y_1, \ldots, y_n) &= \frac{L(\hat{\hat{\boldsymbol{\beta}}}, \hat{\hat{\sigma}}^2)}{L(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2)} = \frac{(2\pi e\hat{\hat{\sigma}}^2)^{-\frac{n}{2}}}{(2\pi e\hat{\sigma}^2)^{-\frac{n}{2}}} = \left(\frac{\hat{\sigma}^2}{\hat{\hat{\sigma}}^2}\right)^{\frac{n}{2}} = \left(\frac{\frac{1}{n}\|\boldsymbol{y} - \hat{\boldsymbol{\mu}}\|^2}{\frac{1}{n}\|\boldsymbol{y} - \hat{\hat{\boldsymbol{\mu}}}\|^2}\right)^{\frac{n}{2}} \\ &= \left(\frac{\|\boldsymbol{y} - \hat{\boldsymbol{\mu}}\|^2}{\|\boldsymbol{y} - \hat{\boldsymbol{\mu}}\|^2 + \|\hat{\boldsymbol{\mu}} - \hat{\hat{\boldsymbol{\mu}}}\|^2}\right)^{\frac{n}{2}} = \left(\frac{1}{1 + \frac{\|\hat{\boldsymbol{\mu}} - \hat{\hat{\boldsymbol{\mu}}}\|^2}{\|\boldsymbol{y} - \hat{\boldsymbol{\mu}}\|^2}}\right)^{\frac{n}{2}} \end{aligned}$$

I regningerne er benyttet, at $\boldsymbol{y} - \hat{\boldsymbol{\mu}} \in L^{\perp}$ og $\hat{\boldsymbol{\mu}} - \hat{\hat{\boldsymbol{\mu}}} \in L$.

Det ses, at $\frac{n-k}{k-k_0} \frac{\|\hat{\boldsymbol{\mu}} - \hat{\hat{\boldsymbol{\mu}}}\|^2}{\|\boldsymbol{y} - \hat{\boldsymbol{\mu}}\|^2}$ udgør en aftagende transformation af q.

Vi har tidligere vist, at

$$\|\boldsymbol{Y} - \hat{\boldsymbol{\mu}}\|^2 = \boldsymbol{U}^{\top}(I - P)\boldsymbol{U} \sim \sigma^2\chi^2(n-k) \; .$$

Tilsvarende kan vises, at

$$\|\hat{\boldsymbol{\mu}} - \hat{\hat{\boldsymbol{\mu}}}\|^2 = \boldsymbol{U}^{\top}(P - P_0)\boldsymbol{U} \sim \sigma^2\chi^2(k-k_0) \; ,$$

jf. opgave 15.

Desuden gælder, at

$$\operatorname{Cov}\left(\boldsymbol{Y} - \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\mu}} - \hat{\hat{\boldsymbol{\mu}}}\right) = O \; ,$$

jf. opgave 16, dvs. $\|\boldsymbol{Y} - \hat{\boldsymbol{\mu}}\|^2$ og $\|\hat{\boldsymbol{\mu}} - \hat{\hat{\boldsymbol{\mu}}}\|^2$ er uafhængige.

Der gælder derfor, at

$$F = \frac{n-k}{k-k_0} \frac{\|\hat{\boldsymbol{\mu}} - \hat{\hat{\boldsymbol{\mu}}}\|^2}{\|\boldsymbol{Y} - \hat{\boldsymbol{\mu}}\|^2} \sim F(k-k_0, n-k) \quad \text{under } H_0.$$

Store værdier af f_{obs} er kritiske for hypotesen.

Kvotienttestet for $H_0 : \boldsymbol{\mu} \in L_0$ fører altså frem til et sædvanligt F-test.

Bemærk om de indgående vektorer, at

- $\hat{\hat{\boldsymbol{\mu}}} \in L_0$,
- $\hat{\boldsymbol{\mu}} - \hat{\hat{\boldsymbol{\mu}}} \in L_1 = L \cap L_0^{\perp}$, det ortogonale komplement til L_0 inden for L ($P_0\hat{\boldsymbol{\mu}} = P_0P\boldsymbol{y} = P_0\boldsymbol{y} = \hat{\hat{\boldsymbol{\mu}}} \Rightarrow \hat{\boldsymbol{\mu}} - \hat{\hat{\boldsymbol{\mu}}} \perp \hat{\hat{\boldsymbol{\mu}}}$),
- $\boldsymbol{y} - \hat{\boldsymbol{\mu}} \in L^{\perp}$.

Det er umiddelbart klart, at $L_0 + L_1 = L$. Da desuden $L_0 \cap L_1 = \{\mathbf{0}\}$, siger vi uddybende, at L er fremkommet som den direkte sum af L_0 og L_1, og vi benytter symbolet $\oplus$ i stedet for $+$, dvs. $L_0 \oplus L_1 = L$. Vi kan også indføre symbolet $\ominus$ ved reglen $L \ominus L_0 = L \cap L_0^{\perp}$, dvs. $L_1 = L \ominus L_0$.

Bemærk, at $L_0 \oplus (L \ominus L_0) \oplus L^{\perp} = R^n$, idet der gælder, at $L_0 \cap (L \ominus L_0) = L_0 \cap L^{\perp} = (L \ominus L_0) \cap L^{\perp} = \{\mathbf{0}\}$. Da rummene endvidere er indbyrdes ortogonale, siger vi, at R^n er fremstillet som den ortogonale direkte sum af de tre underrum.

Lad os betragte $L^{\perp}$ opdelt i r indbyrdes ortogonale underrum, således at

$$
\begin{aligned}
R^n &= L \oplus L^{\perp} & &, \quad \dim L = k \ , \\
&= L \oplus L_1 \oplus \ldots \oplus L_r & &, \quad \dim L_i = k_i \ , \quad \textstyle\sum_i k_i = n - k \ .
\end{aligned}
$$

Der gælder da, at

- $P\boldsymbol{Y}$, $P_1\boldsymbol{Y}, \ldots, P_r\boldsymbol{Y}$ er uafhængige,
- $\|P\boldsymbol{Y} - \boldsymbol{\mu}\|^2 \sim \sigma^2\chi^2(k)$,
- $\|P_i\boldsymbol{Y}\|^2 \sim \sigma^2\chi^2(k_i)\,, \quad i = 1, \ldots, r$,

hvor P_i er projektionsmatricen for projektion på L_i, $i = 1, \ldots, r$.

Beviset henlægges til opgave 17.

4.11 Affine hypoteser

Vi går igen ud fra grundmodellen $\boldsymbol{\mu} \in L$, men betragter nu hypotesen $H_0 : \boldsymbol{\mu} \in \boldsymbol{x_0} + L_0$, dvs. at $\boldsymbol{\mu}$ ligger i et såkaldt affint underrum af L.

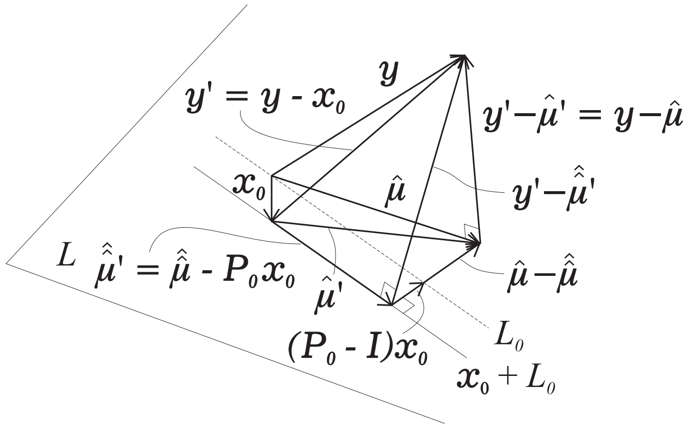
Vi indfører $\boldsymbol{y}' = \boldsymbol{y} - \boldsymbol{x_0}$ og bemærker, at $\mathrm{E}[\boldsymbol{Y}'] = \mathrm{E}[\boldsymbol{Y}] - \boldsymbol{x_0} \Leftrightarrow \boldsymbol{\mu}' = \boldsymbol{\mu} - \boldsymbol{x_0}$. Der gælder åbenbart, at $H_0 : \boldsymbol{\mu} \in \boldsymbol{x_0} + L_0$ er ækvivalent med $H_0 : \boldsymbol{\mu}' \in L_0$.

H_0 kan nu testes ved benyttelse af

$$
\begin{aligned}
F &= \frac{n-k}{k-k_0}\,\frac{\|\hat{\boldsymbol{\mu}}' - \hat{\hat{\boldsymbol{\mu}}}'\|^2}{\|\boldsymbol{Y}' - \hat{\boldsymbol{\mu}}'\|^2} = \frac{n-k}{k-k_0}\,\frac{\|\hat{\boldsymbol{\mu}} - \boldsymbol{x_0} - (\hat{\hat{\boldsymbol{\mu}}} - P_0\boldsymbol{x_0})\|^2}{\|\boldsymbol{Y} - \boldsymbol{x_0} - (\hat{\boldsymbol{\mu}} - \boldsymbol{x_0})\|^2} \\
&= \frac{n-k}{k-k_0}\,\frac{\|\hat{\boldsymbol{\mu}} - \hat{\hat{\boldsymbol{\mu}}} + (P_0 - I)\boldsymbol{x_0}\|^2}{\|\boldsymbol{Y} - \hat{\boldsymbol{\mu}}\|^2} \sim F(k-k_0, n-k) \quad \text{under } H_0.
\end{aligned}
$$

4.12 Opgaver

1. Vis, at en ortogonal dekomposition af en vektor er entydig. Vink: Antag, at der er to forskellige dekompositioner, og betragt deres differens.



Figur 5: Test af affin hypotese $H_0 : \boldsymbol{\mu} \in \boldsymbol{x_0} + L_0$

2. Bestem MK-estimaterne for α og β i den simple lineære regressionsmodel ved
   - projektion på $\mathrm{sp}\{\mathbf{1}, \boldsymbol{t}\}$, hvor $\boldsymbol{t} = \boldsymbol{x} - \bar{x}\mathbf{1}$ (bemærk, at $\mathbf{1} \perp \boldsymbol{t}$),
   - benyttelse af normalligningerne.
3. Kontroller, at $P = X(X^\top X)^{-1}X^\top$ er symmetrisk og idempotent.
4. Redegør for, at fordelingen af $\hat{\boldsymbol{\mu}} \sim N_n(\boldsymbol{\mu}, \sigma^2 P)$ er singulær.
5. Vis, at $I - P$ er en projektionsmatrix, og at $(I - P)\boldsymbol{y}$ er projektionen af $\boldsymbol{y}$ på $L^\perp$.
6. Eftervis, at $\|\boldsymbol{y} - \hat{\boldsymbol{\mu}}\|^2 = \|\boldsymbol{y}\|^2 - \|\hat{\boldsymbol{\mu}}\|^2$.
7. Udled, at $\|\boldsymbol{y}\|^2 - \|\hat{\boldsymbol{\mu}}\|^2 = (\boldsymbol{y} - X\hat{\boldsymbol{\beta}})^\top \boldsymbol{y}$. Vink: Indfør $\hat{\boldsymbol{\mu}} = P\boldsymbol{y}$, foretag en del omskrivninger, og erstat til sidst $P\boldsymbol{y}$ med $X\hat{\boldsymbol{\beta}}$.
8. Vis, at $\mathrm{tr}\, P = k$. Vink: Betragt $P = X(X^\top X)^{-1}X^\top$, og udnyt regnereglen $\mathrm{tr}\,(AB) = \mathrm{tr}\,(BA)$, der gælder, når begge matrixprodukter eksisterer.
9. Vis, at enhver projektionsmatrix netop har egenværdier 0 og 1, og angiv disses multipliciteter. Vink: Vis først, at $\lambda \boldsymbol{x}^\top \boldsymbol{x} = \lambda^2 \boldsymbol{x}^\top \boldsymbol{x}$ ved udnyttelse af $P\boldsymbol{x} = \lambda \boldsymbol{x}$ og $P^2 = P$.
10. Vis, at en projektionsmatrix P har $\mathrm{rang}\, P = \mathrm{tr}\, P$.
11. I afsnit 2.3 blev det påstået, at $S^2 \sim \frac{\sigma^2}{n-2}\chi^2(n-2)$. Argumenter for, at denne påstand er dækket ind gennem udledningerne i afsnit 4.3.
12. Gennemfør regningerne til bestemmelse af ML-estimatet $\hat{\sigma^2}$.

13. Redegør for fordelingerne af $(X^\top X)^{\frac{1}{2}}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ og $(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top X^\top X(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$.

14. Udled i detaljer prædiktionsintervallet for en tænkt fremtidig observation.

15. Vis, at $\|\hat{\boldsymbol{\mu}} - \hat{\hat{\boldsymbol{\mu}}}\|^2 = \boldsymbol{U}^\top(P - P_0)\boldsymbol{U} \sim \chi^2(k - k_0)$.

16. Eftervis, at $\mathrm{Cov}\left(\boldsymbol{Y} - \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\mu}} - \hat{\hat{\boldsymbol{\mu}}}\right) = O$.

17. Lad $L^\perp$ være opdelt i r indbyrdes ortogonale underrum, således at

$$\begin{aligned} R^n &= L \oplus L^\perp & , &\quad \dim L = k \ , \\ &= L \oplus L_1 \oplus \ldots \oplus L_r & , &\quad \dim L_i = k_i \ , \quad \textstyle\sum_i k_i = n - k \ . \end{aligned}$$

Bevis, at

- $P\boldsymbol{Y}$, $P_1\boldsymbol{Y}, \ldots, P_r\boldsymbol{Y}$ er uafhængige
- $\|P\boldsymbol{Y} - \boldsymbol{\mu}\|^2 \sim \sigma^2\chi^2(k)$
- $\|P_i\boldsymbol{Y}\|^2 \sim \sigma^2\chi^2(k_i)\,, \quad i = 1, \ldots, r\ ,$

hvor P_i er projektionsmatricen for projektion på L_i, $i = 1, \ldots, r$.

18. Ved multipel lineær regression forstås normalfordelte observationer af formen

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik}, \ i = 1, \ldots, n \ .$$

Opstil en lineær normal model, der passer til sådanne observationer.
Vink: Designmatricen skal indeholde $k + 1$ søjler med $\mathbf{1}$ som den første.

Hvilke ændringer må der foretages i formlerne udviklet til den generelle lineære model, for at de kan benyttes til multipel lineær regression ?

(Note: I visse fremstillinger kaldes den generelle lineære model for multipel lineær regression.)