Generel lineær model
Antag, $Y \sim \mathcal{N}(\mu, \sigma^2\varepsilon)$. Hvis $\mu \in \Omega_0$, hvor Ω_0 er et lineært underrum af $\mathbb{R}^n$, så kaldes dette en generel lineær model. Dimensionen af Ω_0, $\dim(\Omega_0)$ kadels modellens dimension.

Designmatricen
Antag, at $\Omega_0 = \text{span}(\{x_1, \ldots, x_k\})$, $k \leq n$ og X er en $n \times k$-matrix med søjler $x_1, \ldots, x_k$ med fuld rang k. Så kaldes X designmatricen. Ω_0 er søjlerummet for X og $\mu = X\beta$ for en vektor $\beta \in \mathbb{R}^k$ kaldet parametervektoren.

Bemærk: Da X har fuld rang, så er $\dim(\Omega_0) = k$.

Eksempel (Multipel lineær regression):

$$
\begin{aligned}
y_i &= \beta_0 + \beta X_{i,1} + \ldots + \beta_{k-1} X_{i,k-1} + \varepsilon_i \\
y &= X\beta + \varepsilon
\end{aligned}
$$

hvor

$$
X = \begin{bmatrix} 1 & X_{1,1} & & X_{1,k-1} \\ \vdots & \vdots & & \vdots \\ 1 & X_{n,1} & & X_{n,k-1} \end{bmatrix} \qquad \text{og}
$$

$$
\beta = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_{k-1} \end{bmatrix}
$$

Fuld rang betyder essentielt at de enkelte søjler indeholder forskellig information.

Nrå X ikke har fuld rang, kan vi sommetider estimere en linearkombination af indgangene i β.

Estimerbar linearkombination
En linearkombination $\psi = c^T\beta$ er estimerbar, hvis der eksisterer $a^T y$, så at $E[a^T Y] = c^T\beta$ for alle $\beta \in \mathbb{R}^k$.

Bemærk: Alle $c^T\beta$ er estimerbare hvis X har fuld rang.

Estimation af β
For $\mu = X\beta$, så er MLE for β en løsning normalligningen

$$
X^T \Sigma^{-1} y = X^T \Sigma^{-1} X \hat{\beta}.
$$

Hvis X har fuld rang, så er løsningen entydig og givet ved

$$\hat{\beta} = \left(X^T\Sigma^{-1}X\right)^{-1} X^{-1}\Sigma^{-1}y$$

Bevis 1 (differentiation)
Løs $S(\beta; y) = 0$.

$$\begin{aligned}
S(\beta; y) &= \left(\frac{\partial \mu(\beta)}{\partial \beta}\right)^T S(\mu(\beta); y) \\
&= \left(\frac{\partial}{\partial \beta} X\beta\right)^T \frac{1}{\sigma^2}\Sigma^{-1}(y - X\beta) \\
&= \frac{1}{\sigma^2} X^T\Sigma^{-1}(y - X\beta) \\
&= \frac{1}{\sigma^2}(X^T\Sigma^{-1}y - X^T\Sigma^{-1}X\beta) \\
&= 0 \\
&\Downarrow \\
X^T\Sigma^{-1}y &= X^T\Sigma^{-1}X\beta
\end{aligned}$$

Sidste del af første linje kan findes i formel (3.8). Hvis X har fuld rang k, så har $k \times k$-matricen $X^T\Sigma{-1}X$ også fuld rang. Da er den også inverterbar og den ganges på ovenstående:

$$\hat{\beta} = \left(X^T\Sigma^{-1}X\right)^{-1} X^T\Sigma^{-1}y$$

Bevis 2 (Geometri)
Likelihood:

$$L(\mu; y) = \frac{1}{\sigma^2}\exp -\frac{1}{2\sigma^2}D(y; \mu)$$

Dvs L er maksimeret når $D(y; \mu) = (y - \mu)^T\Sigma^{-1}(y - \mu) = \|y - \mu\|_\Sigma$ er minimeret. $\|y - \mu\|_\Sigma$ er minimeret, når $y - \mu$ er ortogonal på Ω_0. Dette er afstanden fra y til hyperplanen Ω_0.

$$\begin{aligned}
0 &= \delta_\Sigma\left(Xv, y - \hat{\mu}\right), & \forall r \in \mathbb{R}^n \\
&= (Xv)^T\Sigma^{-1}\left(y - X\hat{\beta}\right) \\
&= v^T\left(X^T\Sigma^{-1}y - X^T\Sigma^{-1}X\hat{\beta}\right)
\end{aligned}$$

Da produktet skal være 0 skal v^T være ortogonal på parentesen for alle v. Den eneste vektor der gør dette er nulvektoren. Altså

$$X^T\Sigma^{-1}y = X^T\Sigma^{-1}X\hat{\beta}$$

Egenskaber ved MLE for β
Hvis X har fuld rang, så

$$\hat{\beta} \sim \mathcal{N}_k \left(\beta, \sigma^2 \left(X^T \Sigma^{-1} X\right)^{-1}\right)$$

Bemærk:

$$\begin{aligned} E[\hat{\beta}] &= \beta && \text{central estimator} \\ Var[\hat{\beta}] &= \sigma^2 \left(X^T \Sigma^{-1}\right)^{-1} && \text{efficient estimator} \end{aligned}$$

Bemærk: Da Σ er symmetrisk og positiv definit, så eksisterer en invertibel matrix $A \in \mathbb{R}^{n \times n}$, sådan at $\Sigma = AA^T$. Dvs.

$$\begin{aligned} Y &= X\beta + \varepsilon && \varepsilon \sim \mathcal{N}_n\left(0, \sigma^2 \Sigma\right) \\ &= X\beta + A\tilde{\varepsilon} && \tilde{\varepsilon} \sim \mathcal{N}_n\left(o, \sigma^2 I\right) \\ &\Downarrow \\ A^{-1}Y &= A^{-1}X\beta + \tilde{\varepsilon} \\ &\Downarrow \\ \tilde{Y} &= \tilde{X}\beta 0 \tilde{\varepsilon} \end{aligned}$$

Hvor $\tilde{Y} = A^{-1}Y$ og $\tilde{X} = A^{-1}X\beta$.