APPENDIX: Formulas in STK3100/4100

1) Linear models and least squares

a) Let $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^{\mathrm{T}}$ be a vector of random variables with mean vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^{\mathrm{T}}$ and covariance matrix $\boldsymbol{V} = E\{(\boldsymbol{Y} - \boldsymbol{\mu})(\boldsymbol{Y} - \boldsymbol{\mu})^{\mathrm{T}}\}$. We consider the linear model $\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta}$, where the model matrix $\boldsymbol{X}$ is a $n \times p$ matrix, and assume that $\boldsymbol{V} = \sigma^2 \boldsymbol{I}$. If we observe $\boldsymbol{Y} = \boldsymbol{y} = (y_1, \ldots, y_n)^{\mathrm{T}}$, then the least squares estimate $\widehat{\boldsymbol{\beta}}$ and the fitted values $\widehat{\boldsymbol{\mu}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}$ are obtained by minimizing $\|\boldsymbol{y} - \boldsymbol{\mu}\|^2 = (\boldsymbol{y} - \boldsymbol{\mu})^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{\mu})$.

b) Let $C(\boldsymbol{X})$ denote the model space, i.e. the subspace of $\mathbb{R}^n$ that is spanned by the columns of $\boldsymbol{X}$, and let $\boldsymbol{P}_{\boldsymbol{X}}$ denote the projection matrix onto $C(\boldsymbol{X})$. Then $\widehat{\boldsymbol{\mu}} = \boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{y}$. The projection matrix is symmetric and idempotent (i.e. $\boldsymbol{P}_{\boldsymbol{X}}^2 = \boldsymbol{P}_{\boldsymbol{X}}$), and $\mathrm{rank}(\boldsymbol{P}_{\boldsymbol{X}}) = \mathrm{trace}(\boldsymbol{P}_{\boldsymbol{X}})$.

c) The projection matrix $\boldsymbol{P}_{\boldsymbol{X}}$ is unique, i.e. it depends only on the subspace $C(\boldsymbol{X})$ and not on the choice of basis vectors for the subspace. If $\boldsymbol{X}$ has full rank, we have $\boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}$.

d) For a random vector $\boldsymbol{Y}$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{V}$ and a fixed matrix $\boldsymbol{A}$, we have $E(\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{Y}) = \mathrm{trace}(\boldsymbol{A}\boldsymbol{V}) + \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{\mu}$.

2) Multivariate normal distribution and normal linear models

a) $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^{\mathrm{T}}$ has a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{V}$, written $\boldsymbol{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{V})$, if its joint pdf is given by

$$f(\boldsymbol{y}; \boldsymbol{\mu}, \boldsymbol{V}) = (2\pi)^{-n/2}|\boldsymbol{V}|^{-1/2} \exp\{-(1/2)(\boldsymbol{y} - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{\mu})\}$$

b) Suppose $\boldsymbol{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{V})$ is partitioned as

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \end{pmatrix} \quad \text{with} \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \quad \text{and} \quad \boldsymbol{V} = \begin{pmatrix} \boldsymbol{V}_{11} & \boldsymbol{V}_{12} \\ \boldsymbol{V}_{21} & \boldsymbol{V}_{22} \end{pmatrix}$$

then

$$\boldsymbol{Y}_1 | \boldsymbol{Y}_2 = \boldsymbol{y}_2 \sim N\left(\boldsymbol{\mu}_1 + \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}(\boldsymbol{y}_2 - \boldsymbol{\mu}_2), \boldsymbol{V}_{11} - \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{V}_{21}\right)$$

c) [Cochran's theorem] Assume that $\boldsymbol{Y} \sim N(\boldsymbol{\mu}, \sigma^2\boldsymbol{I})$ and that $\boldsymbol{P}_1, \ldots, \boldsymbol{P}_k$ are projection matrices with $\sum_{i=1}^{k} \boldsymbol{P}_i = \boldsymbol{I}$. Then $\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{P}_i\boldsymbol{Y}$ are independent for $i = 1, \ldots k$, and $\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{P}_i\boldsymbol{Y}/\sigma^2$ has a non-central chi-squared distribution with non-centrality parameter $\lambda_i = \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{P}_i\boldsymbol{\mu}/\sigma^2$ and degrees of freedom equal to the rank of $\boldsymbol{P}_i$.

3) Generalized linear models (GLMs)

a) A random variable Y_i has a distribution in the exponential dispersion family if its pmf/pdf may be written

$$f(y_i; \theta_i, \phi) = \exp\{[y_i\theta_i - b(\theta_i)]/a(\phi) + c(y_i, \phi)\},$$

where θ_i is the natural parameter and ϕ is the dispersion parameter. We have $E(Y_i) = b'(\theta_i)$ and $\mathrm{var}(Y_i) = b''(\theta_i)a(\phi)$.

b) For a GLM we have that $Y_1, \ldots Y_n$ are independent with pmf/pdf from the exponential dispersion family. The linear predictors $\eta_1, \ldots, \eta_n$ are given by $\eta_i = \sum_{j=1}^{p} x_{ij}\beta_j = \boldsymbol{x}_i\boldsymbol{\beta}$, and the expected values $\mu_i = E(Y_i)$ satisfy $g(\mu_i) = \eta_i$ for a strictly increasing and differentiable link function g. For the canonical link function $g(\mu_i) = (b')^{-1}(\mu_i)$ we have $\theta_i = \eta_i$.

c) The likelihood equations for a GLM are given by

$$\sum_{i=1}^{n} \frac{(y_i - \mu_i)x_{ij}}{\mathrm{var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} = 0 \quad \text{for} \quad j = 1, \ldots, p.$$

d) Let $\widehat{\boldsymbol{\beta}}$ be the maximum likelihood (ML) estimator for a GLM. Then

$$\widehat{\boldsymbol{\beta}} \sim N\left(\boldsymbol{\beta}, (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{X})^{-1}\right), \quad \text{approximately}$$

where $\boldsymbol{X}$ is the model matrix and $\boldsymbol{W}$ is the diagonal matrix with elements $w_i = (\partial\mu_i/\partial\eta_i)^2/\mathrm{var}(Y_i)$.

e) Consider a GLM with $a(\phi) = \phi/\omega_i$. Let $\widehat{\mu}_i = b'(\widehat{\theta}_i)$ be the ML estimate of μ_i under the actual model, and let $y_i = b'(\tilde{\theta}_i)$ be the ML estimate of μ_i under the saturated model. Then

$$-2\log\left(\frac{\text{max likelihood for actual model}}{\text{max likelihood for saturated model}}\right) = D(\boldsymbol{y}; \widehat{\boldsymbol{\mu}})/\phi$$

where

$$D(\boldsymbol{y}; \widehat{\boldsymbol{\mu}}) = 2\sum_{i=1}^{n} \omega_i \left[y_i\left(\tilde{\theta}_i - \widehat{\theta}_i\right) - b(\tilde{\theta}_i) + b(\widehat{\theta}_i)\right]$$

is the deviance.

4) Normal and generalized linear mixed models

a) We assume that $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{id})^{\mathrm{T}}$ for $i = 1, \ldots n$ are independent vectors that correspond to d observations from each of n clusters. A normal linear mixed effects model is given by

$$Y_{ij} = \boldsymbol{x}_{ij}\boldsymbol{\beta} + \boldsymbol{z}_{ij}\boldsymbol{u}_i + \epsilon_{ij},$$

where $\boldsymbol{\beta}$ is a $p \times 1$ vector of fixed effects, $\boldsymbol{u}_i \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_u)$ is a $q \times 1$ vector of random effects, and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots \epsilon_{id})^{\mathrm{T}} \sim N(\boldsymbol{0}, \boldsymbol{R})$ is independent of $\boldsymbol{u}_i$. Often one will have $\boldsymbol{R} = \sigma^2\boldsymbol{I}$.

b) For a generalized linear mixed model we assume that the conditional pmf/pdf of Y_{ij} given $\boldsymbol{u}_i \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_u)$ is in the exponential dispersion family, and that for a link function g we have

$$g\left[E(Y_{ij} \mid \boldsymbol{u}_i)\right] = \boldsymbol{x}_{ij}\boldsymbol{\beta} + \boldsymbol{z}_{ij}\boldsymbol{u}_i.$$