**BIOS 667: Longitudinal Data Analysis**
REML - summary

The restricted likelihood (residual likelihood, marginal likelihood) arises in the following setup.

The $n \times 1$ response vector $Y$ is multivariate normal with $\mathrm{E}Y = \mu = X\beta$, where $X$ is $n \times p$, full column rank ($p <= n$), $\mathrm{cov}(Y) = \Sigma(\theta)$, where $\theta$ is a $q \times 1$ vector that determines $\Sigma(\theta)$.

The log-likelihood is

$$l(\beta, \theta; Y) = -\frac{1}{2} \log |\Sigma| - \frac{1}{2}(Y - X\beta)^\top \Sigma^{-1}(Y - X\beta).$$

Note: Additive constants will be dropped from log-likelihoods.

Take any projection matrix, $P$, that projects on the column space of $X$, and form the residuals

$$R = Y - PY = (I - P)Y.$$

The REML log-likelihood is

$$l_{REML}(\theta; R) = -\frac{1}{2} \log |\Sigma| - \frac{1}{2} \log |X^\top \Sigma^{-1} X| - \frac{1}{2} Q(R, \Sigma, X),$$

where

$$Q(R, \Sigma, X) = R^\top \left\{ \Sigma^{-1} - \Sigma^{-1} X (X^\top \Sigma^{-1} X)^{-1} X^\top \Sigma^{-1} \right\} R.$$

The REML likelhood involves only $\Sigma$ (or $\theta$), but not $\beta$. The REML likelihood can not be used for inference about $\beta$. Specifically, likelihood ratio tests for $\beta$ based on the REML likelhood are **not possible**.

The REML likelhood is invariant to the choice of $P$. That is, even though different choices of $P$ lead to different residuals $R$, they lead to the same REML likelihood.

The REML likelihood is a genuine likelihood and can be used like any ordinary likelihood for inference about $\theta$. This includes REML estimation of $\theta$, likelihood ratio, score and Wald tests, Fisher information, etc. Maxmimization of $l_{REML}$ is typically an iterative computation using Fisher scoring or other methods.

Once $\hat{\theta}$, and hence $\hat{\Sigma}$, is available from REML, an estimate of $\beta$ can be obtained by a WLS-like computation as $\hat{\beta} = (X^\top W X)^{-1} X^\top W Y$ where $W = \{\Sigma(\hat{\theta})\}^{-1}$. This is confusingly called the "REML estimator of $\beta$" even though the REML likelihood actually does not involve $\beta$ at all (we can't maximize the REML likelihood over $\beta$).

The model-based estimator of $\mathrm{cov}(\hat{\beta})$ is $(X^\top W X)^{-1}$. This matrix can be used for inference (confidence intervals, Wald-type tests) about $\beta$.

Unlike WLS (where $W$ is a non-random matrix), $\hat{\beta}$ here is generally biased. The reason is that $W$, being a function of $Y$, is random and does not factor out of expectations as it does in WLS.

In longitudinal and clustered data setups, the matrix $\Sigma$ is block-diagonal with blocks $\Sigma_i$, and $X$ is similary partitioned into $\{X_i\}$. In this case, some of the above expressions can be written as summations. For example, $X^\top \Sigma^{-1} X = \sum_{i=1}^{K} X_i^\top \Sigma_i^{-1} X_i$, where $K$ is the number of subjects or clusters.