Statistical Inference

Inference can be based on the ML estimator $\hat{\beta}$ and its estimated covariance matrix

$$\hat{\text{Cov}}(\hat{\beta}) = \left\{ \sum_{i=1}^{N} \left(X_i'\hat{\Sigma}_i^{-1}X_i\right) \right\}^{-1}$$

For a single linear combination $L\beta$:

- approximate $100(1 - \alpha)$% confidence interval

$$L\hat{\beta} \pm Z_{\alpha/2}\sqrt{L\hat{\text{Cov}}(\hat{\beta})L'}$$

- Test $H_0 : L\beta = 0$ with the $N(0,1)$ Wald statistic

$$Z = \frac{L\hat{\beta}}{\sqrt{L\hat{\text{Cov}}(\hat{\beta})L'}}$$
For \textit{more than one} linear combination $L \beta$, where $L$ is $r \times p$, test $H_0 : L \beta = 0$ with the Wald statistic

$$W^2 = (L \hat{\beta})' \{L \tilde{\text{Cov}}(\beta)L\}'^{-1} (L \hat{\beta})$$

which under $H_0$ is approximately $\chi^2$ with $r$ degrees of freedom.

Note: if $r = 1$, $W^2 = Z^2$.

Example: to test $H_0 : \beta_1 = \beta_2 = \beta_3$, write $H_0$ as $L \beta = 0$ with

$$L = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 1 & -1 \end{pmatrix}$$

The rows of $L$ can be any two contrasts: $W^2$ has the same value for any choice.
A (preferred) alternative to the Wald test: the *likelihood ratio* test. To test $H_0 : L\beta = 0$:

- fit the model with this constraint (the “reduced” model), with maximized log likelihood $\hat{\ell}_{\text{reduced}}$;

- compare with the model fitted without the constraint (the “full” model), with maximized log likelihood $\hat{\ell}_{\text{full}}$.

Then under $H_0$

$$G^2 = 2(\hat{\ell}_{\text{full}} - \hat{\ell}_{\text{reduced}})$$

is approximately $\chi^2$ with $r$ degrees of freedom.
Likelihood ratio confidence interval (confidence region if $r > 1$) for $L\beta$:

- the profile likelihood $l_p$ is
  \[ l_p(x; L) = \max_{L\beta=x} l(\beta) \]

- an approximate $100(1-\alpha)%$ confidence interval is
  \[ \left\{ x : 2 \times \left\{ \hat{l}_{\text{full}} - l_p(x; L) \right\} \right\} \leq \chi^2_r(\alpha) \]

This is exactly the set of values $x$ for which $H_0 : L\beta = x$ is not rejected at level $\alpha$. 
Denominator degrees of freedom

- These normal and $\chi^2$ distributions are \textit{large sample approximations}.

- In simple cases (balanced, etc.), we can allow for estimating variances and covariances by using \textit{exact} $t$– and $F$–distributions.

- In some other cases, we can use approximations due to Satterthwaite and Kenward and Roger to find \textit{approximate} $t$– and $F$–distributions, typically with approximated degrees of freedom.
Restricted Maximum Likelihood (REML)

- Recall the covariance parameters $\theta$:

$$\text{Cov}(Y_i) = \Sigma_i = \Sigma_i(\theta)$$

- Maximum likelihood estimation of $\theta$ typically leads to biased estimators.

E.g., for a single random sample from $N(\mu, \sigma^2)$,

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^{N} (y_i - \bar{y})^2$$

with expected value $\sigma^2 \times (1 - 1/N)$. 
• We transform the response \( Y \) and partition it as \( (Y_1^*, Y_2^*) \) so that the likelihood factorizes:

\[
l_1(Y_1^*, \theta, \beta) \times l_2(Y_2^*, \theta)
\]

where in some sense, the first factor contains no information about \( \theta \) in the absence of information about \( \beta \).

• We estimate \( \beta \) by maximizing \( l_1 \) and \( \theta \) by maximizing \( l_2 \).

• The resulting \( \hat{\theta} \) gives less biased covariance estimates

\[
\hat{\Sigma}_i = \Sigma_i(\hat{\theta})
\]

E.g., for a single random sample, \( \hat{\sigma}^2_{\text{REML}} = s^2 \).