The idea of REML: (conceptual, not necessarily the computational algorithm used)

(1) In a fixed effect least square procedure (regression, ANOVA), the likelihood is maximized by setting $\sigma^2$ equal to the sum of squared residuals divided by $n$, not by degrees of freedom.

Example: Simple random sample N( , )

$Y = 5, 8, 2, 1$ mean is 4, residuals 1, 4, -2, -3 MLE of $\sigma^2$ is $30/4=7.5$

(2) Dividing by $n$ would be fine IF you knew the true mean or in general the true beta in $X$ beta. The fitted mean (estimated beta) fits the data better than the true one.

(3) The trick: Extract a set of numbers from the residuals that have the same variance (or variance component) as the original data but have known mean 0.

Example: $\begin{pmatrix}
.5 & .5 & .5 & .5 \\
- .5 & - .5 & .5 & .5 \\
- .5 & .5 & - .5 & .5 \\
.5 & - .5 & - .5 & .5
\end{pmatrix}$

$\begin{pmatrix}
5 \\
8 \\
2 \\
1
\end{pmatrix}$

Note: For residuals $\begin{pmatrix}
.5 & .5 & .5 & .5 \\
- .5 & - .5 & .5 & .5 \\
- .5 & .5 & - .5 & .5 \\
.5 & - .5 & - .5 & .5
\end{pmatrix}$

$\begin{pmatrix}
1 \\
4 \\
-1 \\
-3
\end{pmatrix}$

First element ALWAYS 0 (not random).

Other elements ALWAYS the same as before.

Three random elements have the SAME sum of squares

SSE = 25+1+4 = 1+16+4+9 as the 4 residuals.

Three random elements have mean 0

MLE based on three random elements is SSE/3 as now $n$ is 3.

This is REML – less biased than ML, unbiased in this particular example.

(A) Run regression and get residuals. Residuals = $(I - X(X'X)^{-1}X')Y = PY = PR$, that is, annihilate the fixed effects.

(B) Maximize the likelihood of linear combinations of these residuals that have mean 0 and behave like the T example above.

This extends to the general kinds of covariance structures handled by MIXED.