Partitioning of Total Sum of Squares
and Coefficient of Determination ($R^2$)

\[ y \]
\[ \text{SRF: } \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \]
\[ \overline{y} \]
\[ 0 \quad X_i \quad X \]

Total deviation from mean
\[ y_i - \overline{y} = (1) + (2) \]

Deviation due to regression
\[ \hat{y}_i - \overline{y} = (1) \]

Deviation due to error
\[ y_i - \hat{y}_i \]

\[ \therefore \text{Total deviation} = \text{Deviation due to regression} + \text{Deviation due to error} \]

\[ y_i - \overline{y} = (\hat{y}_i - \overline{y}) + (y_i - \hat{y}_i) \]
As it turns out

\[
\sum_{i} (y_i - \bar{y})^2 = \sum_{i} (y_i - \hat{y}_i + \hat{y}_i - \bar{y})^2
\]

\[
= \sum_{i} (y_i - \hat{y}_i)^2 + 2 \sum_{i} (y_i - \hat{y}_i)(\hat{y}_i - \bar{y})^2
\]

\[
+ \sum_{i} (\hat{y}_i - \bar{y})^2
\]

\[
= \sum_{i} (y_i - \hat{y}_i)^2 + \sum_{i} (\hat{y}_i - \bar{y})^2
\]

( because it can be shown that \( \sum_{i} (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = 0 \) )

\[
sst = ssr + sse
\]

Total sum of squares = sum of squared residuals + sum of squares explained.

The coefficient of determination \( R^2 \) is defined as

\[
R^2 = \frac{sse}{sst} = \frac{\sum_{i} (\hat{y}_i - \bar{y})^2}{\sum_{i} (y_i - \bar{y})^2}
\]

and \( 0 \leq R^2 \leq 1 \).
$R^2$ is interpreted as the percent (indecidual equivalent form) of the variation in $y$ explained by the explanatory variables of a regression.

**Analysis of Variance (ANOVA)**

**Table**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explained</td>
<td>$k-1$</td>
<td>SSE</td>
<td>$\frac{SSE}{k-1}$</td>
<td>$rac{MS}{MSSR}$</td>
</tr>
<tr>
<td>Error</td>
<td>$N-k$</td>
<td>SSR</td>
<td>$\frac{SSR}{N-k}$</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>$N-1$</td>
<td>SST</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The null hypothesis of interest for the ANOVA table is

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_{k-1} = 0$$

$$H_1 : \text{not } H_0.$$

This is called the **overall test of significance**
of a multiple regression. The corresponding
F-test is called the overall-F-test.
This F-test (in repeated samples) follows
an F distribution with k-1 numerator
degrees of freedom and N-k denominator
degrees of freedom under the truth of
the null hypothesis (H₀). If the
probability value of the F-statistic
of the ANOVA table is greater than
α (usually α = 0.05) then we accept the
null hypothesis that all of the explanatory
variables are jointly insignificant and the
sample mean provides an adequate description
of the variation in the dependent variable.
If the F-statistic's p-value is less
than $\alpha$, we reject the null hypothesis and accept the alternative hypothesis that one or more of the proposed explanatory variables provides significant explanatory power in describing the variation in the dependent variable $y$. 