Test for Lack of Fit

Lack of Fit

Degrees of freedom

Example: Growth rate data

Contributors

Lack of Fit

When we have repeated measurements for different values of the predictor variables \(X\), it is possible to test whether a linear model fits the data.

Suppose that we have data that can be expressed in the form:

\[
\{(X_j, Y_{ij}) : i = 1, \ldots, n_j; j = 1, \ldots, c\} \] where \(c > 2\).

Assume that the data come from the model:

\[
Y_{ij} = \mu_j + \epsilon_{ij}, \quad i = 1, \ldots, n_j; \quad j = 1, \ldots, c (1)
\]

The null hypothesis in which the linear model holds is: \(H_0: \mu_j = \beta_0 + \beta_1 X_j\), for all \(j = 1, \ldots, c\).

Here (1) is the full model and the model specified by \(H_0\) is the reduced model. We follow the usual procedure for the ANOVA, by computing the sum of squares due to errors for the full and reduced models.

Let \(\bar{Y} = \frac{1}{n_j} \sum_{i = 1}^{n_j} Y_{ij}\), and \(\bar{Y} = \frac{1}{c} \sum_{j=1}^{c} n_j \bar{Y}_j\).

Let \(\bar{Y} = \frac{n}{n_j} \sum_{i = 1}^{n}(n_j) Y_{ij}\), and \(\bar{Y} = \frac{c}{c} \sum_{j=1}^{c} n_j \bar{Y}_j\).


Updated: Thu, 02 Jul 2020 05:04:41 GMT
Powered by
\(\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n_j}Y_{ij}\), where \(n = \sum_{j=1}^{c}n_{j}\).

\(\text{SSTO} = \sum_{j=1}^{c}\sum_{i=1}^{n_j}(Y_{ij} - \bar{Y})^2\) and

\(\text{SSPE} = \text{SSE}_{\text{full}} = \sum_{j=1}^{c}\sum_{i=1}^{n_j}(Y_{ij} - \bar{Y}_{j})^2 = \sum_{j=1}^{c}\sum_{i=1}^{n_j}Y_{ij}^2 - \text{SSR}\)

\(\text{SS}_{\text{red}} = \text{SSE} - \text{SSLF}\)

\(\text{SSLF} = \text{SS}_{\text{red}} - \text{SS}_{\text{full}}\)

\(d.f.(\text{SSPE}) = n - c; d.f.(\text{SSLF}) = d.f.(\text{SSE}_{\text{red}}) - d.f.(\text{SSPE}) = (n - 2) - (n - c) = c - 2\).

\[
\begin{array}{|c|c|c|c|c|}
\hline
\text{Source} & \text{d.f.} & \text{SS} & \text{MS = SS/d.f.} & \text{F-statistic} \\
\hline
\text{Regression} & 1 & \text{SSR} & \text{MSR} & \text{MSR/MSE} \\
\hline
\text{Error} & n-2 & \text{SSE} = \text{SSLF} + \text{SSPE} & \text{MSE} & \\
\hline
\text{Lack of fit} & c-2 & \text{SSLF} & \text{MSLF} & \text{MSLF/MSPE} \\
\hline
\text{Pure error} & n-c & \text{SSPE} & \text{MSPE} & \\
\hline
\text{Total} & n-1 & \text{SSTO} = \text{SSR} + \text{SSLF} + \text{SSPE} & & \\
\hline
\end{array}
\]

Reject \(H_0: (\mu_{j} = \beta_0 + \beta_1X_{j} \text{ for all } j)\) at level \(\alpha\) if \(F^*_{LF} = \frac{\text{MSLF}}{\text{MSPE}} > F(1 - \alpha; c - 2, n - c)\).

\[\text{Example: Growth rate data}\]

In the following example, data are available on the effect of dietary supplement on the growth rates of rats. Here \(X = \) dose of dietary supplement and \(Y = \) growth rate. The following table presents the data in a form suitable for the analysis.

\[
\begin{array}{|l|l|l|l|l|l|l|}
\hline
\text{(j = 1)} & \text{(j = 2)} & \text{(j = 3)} & \text{(j = 4)} & \text{(j = 5)} & \text{(j = 6)} \\
\text{(X\_1 = 10)} & \text{(X\_2 = 15)} & \text{(X\_3 = 20)} & \text{(X\_4 = 25)} & \text{(X\_5 = 30)} & \text{(X\_6 = 35)} \\
\text{(n\_1 = 2)} & \text{(n\_2 = 2)} & \text{(n\_3 = 2)} & \text{(n\_4 = 3)} & \text{(n\_5 = 1)} & \text{(n\_6 = 2)} \\
\hline
\text{(Y\_11 = 73)} & \text{(Y\_12 = 85)} & \text{(Y\_13 = 90)} & \text{(Y\_14 = 87)} & \text{(Y\_15 = 75)} & \text{(Y\_16 = 65)} \\
\text{(Y\_21 = 78)} & \text{(Y\_22 = 88)} & \text{(Y\_23 = 91)} & \text{(Y\_24 = 86)} & \text{(Y\_25 = 63)} & \\
\hline
\end{array}
\]

\( \bar{Y}_1 = 75.5 \)
\( \bar{Y}_2 = 86.5 \)
\( \bar{Y}_3 = 90.5 \)
\( \bar{Y}_4 = 88 \)
\( \bar{Y}_5 = 75 \)
\( \bar{Y}_6 = 64 \)

So, for this data, \( c = 6, n = \sum_{j=1}^{c} n_{j} = 12 \).

\( \text{SSTO} = \sum_{j} \sum_{i} (Y_{ij} - \bar{Y})^2 = 1096.00 \)
\( \text{SSPE} = \sum_{j} \sum_{i} Y_{ij}^2 - \sum_{j} n_{j} \bar{Y}_{j}^2 = 79828 - 79704.5 = 33.50 \)
\( \text{SSE}_{\text{red}} = \sum_{i} \sum_{j} (Y_{ij} - \beta_0 - \beta_1 X_{j})^2 = 891.73 \) (Note: \( \beta_0 = 92.003, \beta_1 = -0.498 \))
\( \text{SSLF} = \text{SSE}_{\text{red}} - \text{SSPE} = 891.73 - 33.50 = 858.23 \)

ANOVA Table:

<table>
<thead>
<tr>
<th>Source</th>
<th>d.f.</th>
<th>SS</th>
<th>MS</th>
<th>F*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>204.27</td>
<td>204.27</td>
<td>2.29</td>
</tr>
<tr>
<td>Error</td>
<td>10</td>
<td>891.73</td>
<td>89.173</td>
<td></td>
</tr>
<tr>
<td>Lack of fit</td>
<td>4</td>
<td>858.23</td>
<td>214.56</td>
<td></td>
</tr>
<tr>
<td>Pure error</td>
<td>6</td>
<td>33.50</td>
<td>5.583</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>11</td>
<td>1096.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( F(0.95;4,6) = 4.534 \). Since \( F_{\text{LF}} = 38.43 > 4.534 \), reject \( H_{(0)} : \mu_j = \beta_0 + \beta_1 X_{j} \) for all \( j \) at 5% level of significance.

Also, if you are testing \( H_{(0)^*} : \beta_1 = 0 \) against \( H_{(1)^*} : \beta_1 \neq 0 \), assuming that the linear model holds, as in the usual ANOVA for linear regression, then the corresponding test statistic \( F^* = \frac{\text{MSR}}{\text{MSE}} = 2.29 \). Which is less than \( F(0.95;1,n-2) = F(0.95;1,10) = 4.964 \). So, if one assumes that linear model holds then the test cannot reject \( H_{(0)^*} \) at 5% level of significance.
Contributors

• Joy Wei
• Debashis Paul