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:

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

where \(c > 2\).

Assume that the data come from the model:

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

The null hypothesis in which the linear model holds is: \(H_0: \mu_j = \beta_0 + \beta_1X_j\), for all \(j = 1, ..., c\).

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

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

\(\text{SSE}_{\text{red}} = \text{SSLF} + \text{SSPE}\)

\(\text{SSLF} = \sum_{j=1}^{c}\sum_{i=1}^{n_j}(Y_{ij} - \beta_{0} - \beta_{1}X_{j})^2\)

Degrees of freedom

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

<table>
<thead>
<tr>
<th>Source</th>
<th>d.f.</th>
<th>SS</th>
<th>MS = SS/d.f.</th>
<th>F-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>SSR</td>
<td>MSR</td>
<td>MSR/MSE</td>
</tr>
<tr>
<td>Error</td>
<td>n-2</td>
<td>SSE=SSLF+SSPE</td>
<td>MSE</td>
<td></td>
</tr>
<tr>
<td>Lack of fit</td>
<td>c-2</td>
<td>SSLF</td>
<td>MSLF</td>
<td>MSLF/MSPE</td>
</tr>
<tr>
<td>Pure error</td>
<td>n-c</td>
<td>SSPE</td>
<td>MSPE</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>n-1</td>
<td>SSTO=SSR+SSLF+SSPE</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Reject \(\text{H}_0: (\mu_{j} = \beta_{0} + \beta_{1}X_{j})\) for all \(j\)) at level \(\alpha\) if \((F_{\text{LF}}^*) = \frac{\text{MSLF}}{\text{MSPE}} > F(1 - \alpha; c - 2, n - c)).\)

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.

<table>
<thead>
<tr>
<th>(j = 1)</th>
<th>(j = 2)</th>
<th>(j = 3)</th>
<th>(j = 4)</th>
<th>(j = 5)</th>
<th>(j = 6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(X_{{1}} = 10)</td>
<td>(X_{{2}} = 15)</td>
<td>(X_{{3}} = 20)</td>
<td>(X_{{4}} = 25)</td>
<td>(X_{{5}} = 30)</td>
<td>(X_{{6}} = 35)</td>
</tr>
<tr>
<td>(n_{{1}} = 2)</td>
<td>(n_{{2}} = 2)</td>
<td>(n_{{3}} = 2)</td>
<td>(n_{{4}} = 3)</td>
<td>(n_{{5}} = 1)</td>
<td>(n_{{6}} = 2)</td>
</tr>
<tr>
<td>(Y_{{1}_1} = 73)</td>
<td>(Y_{{2}_1} = 85)</td>
<td>(Y_{{3}_1} = 90)</td>
<td>(Y_{{4}_1} = 87)</td>
<td>(Y_{{5}_1} = 86)</td>
<td>(Y_{{6}_1} = 83)</td>
</tr>
<tr>
<td>(Y_{{1}_2} = 78)</td>
<td>(Y_{{2}_2} = 88)</td>
<td>(Y_{{3}_2} = 91)</td>
<td>(Y_{{4}_2} = 89)</td>
<td>(Y_{{5}_2} = 85)</td>
<td>(Y_{{6}_2} = 82)</td>
</tr>
</tbody>
</table>

Updated: Thu, 23 Jul 2020 03:06:09 GMT
Powered by 2
\( \bar{Y}_{1} = 75.5 \)
\( \bar{Y}_{2} = 86.5 \)
\( \bar{Y}_{3} = 90.5 \)
\( \bar{Y}_{4} = 88 \)
\( \bar{Y}_{5} = 75 \)
\( \bar{Y}_{6} = 64 \)

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: \( \beta_{0} = 92.003, \beta_{1} = -0.498 \))

\( \text{SSLF} = \text{SSE}_{\text{red}} - \text{SSPE} = 891.73 - 33.50 = 858.23 \)

\( \text{d.f.}(\text{SSPE}) = n - c = 6 \)

\( \text{d.f.}(\text{SSLF}) = c - 2 = 4 \)

\( \frac{\text{MSLF}}{\text{MSPE}} = 38.43 \)

\( \text{MSPE} = \frac{\text{SSPE}}{n - c} = 5.5833 \)

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>\frac{\text{MSR}}{\text{MSE}} = 2.29</td>
</tr>
<tr>
<td>Error</td>
<td>10</td>
<td>891.73</td>
<td>89.173</td>
<td>\frac{\text{MSR}}{\text{MSE}} = 2.29</td>
</tr>
<tr>
<td>Lack of fit</td>
<td>4</td>
<td>858.23</td>
<td>214.56</td>
<td>\frac{\text{MSLF}}{\text{MSPE}} = 38.43</td>
</tr>
<tr>
<td>Pure error</td>
<td>6</td>
<td>33.50</td>
<td>5.583</td>
<td>\frac{\text{MSLF}}{\text{MSPE}} = 38.43</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 \( F_{\text{LF}}^\star = 38.43 > 4.534 \), 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 \( F^\star = \frac{\text{MSR}}{\text{MSE}} = 2.29 \). 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
- Debasish Paul