Math 308: Multiple Linear Regression—A General Overview

**Multiple linear regression: Population model and parameters**
The population model for multiple linear regression with p predictors is:

\[ Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \ldots + \beta_{p-1} X_{i,p-1} + \epsilon_i, \]

where \( i \) is the \( i^{th} \) element of the population;

\( \epsilon_i \sim N(0, \sigma^2) \) and \( \sigma^2 \) is constant across all values of \( X_{i,j}, i=1,.., p-1; j=1,2,3,\ldots; \).

Note that these are the assumptions for simple linear regression. Also, as

**Estimates of the model parameters**
As in simple linear regression, the estimates of the coefficients \( \hat{\beta}_j \) are least-square estimates which minimize the sum of the squared residuals (or errors).

**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>Regression</td>
<td>( p - 1 )</td>
<td>SSR</td>
<td>MSR = SSR / (( p - 1 ))</td>
<td>MSR / MSE</td>
</tr>
<tr>
<td>Error</td>
<td>( n - p )</td>
<td>SSE</td>
<td>MSE = SSE / (( n - p ))</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>( n - 1 )</td>
<td>SSTO</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Coefficient of determination, \( R^2 \) and adjusted \( R^2 \)**

\[ R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \]

will stay the same or increase as we add predictor variables. For this reason it is not useful for deciding if a new variable should be added.

\[ \text{Adjusted } R^2 = 1 - \frac{n - 1}{n - p} (1 - R^2) \]

is used for model-building decisions. It has no clear interpretation.

Example: A mobile ad hoc computer network comprises multiple computer that move within a network area. When a receiving node is out of range, the message must be sent to a closer node according to a routing protocol. In this regression Goodput (the percent of messages that are successfully delivered) is a function of average node speed, average pause time.

\[ \text{Goodput} = \beta_0 + \beta_1 \text{speed} + \beta_2 \text{pause} + \beta_3 \text{speed} \times \text{pause} + \beta_4 \text{speed}^2 + \beta_5 \text{pause}^2 + \epsilon \]
We test the overall regression model with an F test. We test the individual predictors (such as speed) with t-tests.

References