Multiple Regressions:

First Order model with two predictors:

\[ Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i. \]

Our assumptions are still the same:

\[ E(\varepsilon_i) = 0 \]
\[ \text{Var} (\varepsilon_i)= \sigma^2 \]

Interpretation of coefficients:

\( \beta_0 \): value of \( Y \) when both \( x_1=0 \) and \( x_2=0 \). (Not much physical meaning beyond this)

\( \beta_1 \): change in response for one unit change in \( X_1 \) given that \( X_2 \) is kept fixed.
   Partial slope of \( X_1 \)

\( \beta_2 \): change in response for one unit change in \( X_2 \) given that \( X_1 \) is kept fixed.
   Partial slope of \( X_2 \)
A first order model

A first order model represents a plane or an unwarped surface.
Now let $x_3 = x_1^2$

As we get further from independence of our explanatory variables we get into the surfaces getting warped.
Let $x_4 = x_1 \times x_2$ (Interaction model)

The SAS System

Let $x_5 = x_2 \times x_2$

The SAS System
First order model in more than two predictors:

\[ Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_{(p-1)} x_{(p-1)i} + \varepsilon_i. \]

\[ E(Y_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_{(p-1)} x_{(p-1)i} \]

This will represent a hyper-plane or a plane in \((p-1)\) dimensions.

Cases we will look at:

1. \((p-1)\) predictor variables
2. Qualitative predictor variables
3. Polynomial regression
4. Transformed variables
5. Interactions
6. Combinations
General Linear Model Regression

For the model

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_{(p-1)} x_{(p-1)i} + \varepsilon_i.$$ 

$\beta_0, \beta_1, \beta_2, \ldots, \beta_{(p-1)}$ are the partial slopes or parameters

$x_{1i}, x_{2i}, \ldots, x_{p-1i}$ are the known constants

$\varepsilon_i$ are independent error terms. $\varepsilon_i \sim N(0, \sigma^2)$

KEEP IN MIND that $x_{1i}, x_{2i}, \ldots, x_{p-1i}$ DO NOT need to represent different predictor variables.

Alternative notation:

$$Y_i = \sum_{j=0}^{p} \beta_j x_{ji} + \varepsilon_i$$

here $x_{0i} = 1$.

$(p-1)$ predictor variables

When $x_{1i}, x_{2i}, \ldots, x_{p-1i}$ represent $(p-1)$ different predictor variables.

This represents a first order model in which there are no interaction effects.
Qualitative predictor variables

Often predictor variables will be qualitative in nature. Examples: color, sex, opinions etc. One uses indicator (dummy) variables in these situations.

\[ X_{2i} = \begin{cases} 
0 & \text{if male} \\
1 & \text{if female} 
\end{cases} \]

In general we need \((c-1)\) predictor variables to account for a qualitative variables with ‘c’ categories.

Variable: agreement status
Categories: Strongly agree
Agree
Disagree

Define: \(X_{1i} = \begin{cases} 
1 & \text{if strongly agree} \\
0 & \text{otherwise} 
\end{cases} \)

Define \(X_{2i} = \begin{cases} 
1 & \text{if agree} \\
0 & \text{otherwise} 
\end{cases} \)

Let \(X_{3i}\) indicate age of the person and \(Y_i\) the numerical response.

Then the model is:
\[ Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_3 + \epsilon_i \]

\[ E(Y_i) = \begin{cases} 
\beta_0 + \beta_1 x_{1i} + \beta_3 x_3 & \text{if they strongly agree} \\
\beta_0 + \beta_2 x_{1i} + \beta_3 x_3 & \text{if they agree} \\
\beta_0 + \beta_3 x_3 & \text{if they disagree} 
\end{cases} \]
Polynomial Regression:

Model of form:
\[ Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{1i}^2 + \beta_3 x_{3i} + \varepsilon_i \]

Define \( x_{2i} = x_{1i}^2 \)

Model:
\[ Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i \]
Transformed variables:

Model can be of form:
Log($Y_i$) = $\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \epsilon_i$

Define $Y_i^* = \log(Y_i)$.

Model;
$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \epsilon_i$
Interaction effects:

Model:

\[ Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i} + \epsilon_i \]

Define: \( x_{3i} = x_{1i} x_{2i} \)

Model:

\[ Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \epsilon_i \]
Combination effects:

Model:
\[ Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i} + \beta_4 x_{1i}^2 + \varepsilon_i \]

Define: \( x_{3i} = x_{1i} x_{2i} \)
And
\( x_{4i} = x_{1i}^2 \)

Model:
\[ Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \varepsilon_i \]

LINEAR indicates LINEAR in the parameters.
Multiple Regressions

The following topics in multiple regressions are interpreted in the same way as in linear regression.
1. Tests on individual $\beta$’s
2. ANOVA Table, F-test
3. Coefficient of determination
4. Confidence intervals for mean
5. Prediction intervals
6. Working-Hotelling confidence regions for the surface
Diagnostics and Remedial measures

For x
1. Dot-plot
2. Stem and leaf Plot
3. Histograms

For residuals
1. Residual Vs. Predicted
2. Residual Vs. X’s
3. Normal Probability Plot

Tests:
1. Correlation test for normality
2. Modified Levene’s test or Brown Forsythe Test
3. Breusch-Pagan Test
4. Lack of fit test
Special Topics in multiple Regression:

Extra Sum of Squares:

Idea: Suppose we are trying to fit the model

\[ Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon \]

with two predictors \( X_1 \) and \( X_2 \).

We may want to find out how much reduction is seen in SSE when we add \( X_2 \) to our model when \( X_1 \) is already in the model.

What is the marginal reduction of the SSE due to the addition of the 2\textsuperscript{nd} predictor variable, when the first one is already in the model?

Notation: \( \text{SSR}(X_2|X_1) = \text{SSE}(X_1) - \text{SSE}(X_1, X_2) \)

\( \text{SSE}(X_1) \) is the sum of squares error for the model

\[ Y = \beta_0 + \beta_1 X_1 + \epsilon \]

\( \text{SSE}(X_1, X_2) \) is the sum of squares error for the model

\[ Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon \]
Alternatively we can think of it from the point of view of Sums of Square Regression.
SSR($X_2|X_1$) = SSR($X_1,X_2$) – SSR($X_1$).

Some results:
SSR($X_2|X_1$) = SSE($X_1$) – SSE($X_1,X_2$)
OR
SSR($X_2|X_1$) = SSR($X_1,X_2$) – SSR($X_1$).

SSR($X_1|X_2$) = SSE($X_2$) – SSE($X_1,X_2$)
OR
SSR($X_1|X_2$) = SSR($X_1,X_2$) – SSR($X_2$).

SSR($X_3|X_1,X_2$) = SSE($X_1,X_2$) – SSE($X_1,X_2,X_3$)
OR
SSR($X_3|X_1,X_2$) = SSR($X_1,X_2,X_3$) – SSR($X_1,X_2$).

SSR($X_3, X_2|X_1$) = SSE($X_1$) – SSE($X_1,X_2,X_3$)
OR
SSR($X_3, X_2|X_1$) = SSR($X_1,X_2,X_3$) – SSR($X_1$).

ANOVA TABLE:

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>SS</th>
<th>DF</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>REGRESSION</td>
<td>SSR($X_1,X_2,X_3$)</td>
<td>3</td>
<td>MSR($X_1,X_2,X_3$)</td>
</tr>
<tr>
<td>$X_1$</td>
<td>SSR($X_1$)</td>
<td>1</td>
<td>MSR($X_1$)</td>
</tr>
<tr>
<td>$X_2</td>
<td>X_1$</td>
<td>SSR($X_2</td>
<td>X_1$)</td>
</tr>
<tr>
<td>$X_3</td>
<td>X_1,X_2$</td>
<td>SSR($X_3</td>
<td>X_1,X_2$)</td>
</tr>
<tr>
<td>ERROR</td>
<td>SSE($X_1,X_2,X_3$)</td>
<td>N-4</td>
<td>MSE($X_1,X_2,X_3$)</td>
</tr>
<tr>
<td>TOTAL</td>
<td>SSTO</td>
<td>N-1</td>
<td></td>
</tr>
</tbody>
</table>
RESULTS:
SSR(X1,X2,X3) = SSR(X1) + SSR(X2|X1) + SSR(X3|X1,X2)

OR
SSR(X1,X2,X3) = SSR(X2) + SSR(X1|X2) + SSR(X3|X1,X2)

OR
SSR(X1,X2,X3) = SSR(X1) + SSR(X2,X3|X1)
TESTING USING EXTRA SUMS OF SQUARES:

To test H0: $\beta_k = 0$

Test statistic: $t^* = b_k / s(b_k)$

Using general Linear Testing: H0: $\beta_3 = 0$ (say) for X1, X2, X3 in the model

Full Model:
$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$

$SSE(\text{Full}) = SSE(X_1, X_2, X_3)$

Reduced Model
$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$

$SSE(\text{Reduced}) = SSE(X_1, X_2)$
Summary of Tests concerning β’s

1. Test whether all β_k =0
   i.e H0: β_1=β_2=... =β_{p-1}

Full Model:
\[ Y = \beta_0 + x_1 \beta_1 + x_2 \beta_2 + \ldots + x_{p-1} \beta_{p-1} \]
\[ \text{SSE(Full)} = \text{SSE}(X_1, \ldots, X_{p-1}) = \text{SSE} \]

Reduced Model:
\[ Y = \beta_0 \]
\[ \text{SSE(Reduced)} = \text{SSTO} \]
\[ F = \frac{[\text{SSTO}-\text{SSE}]/(p-1)]}{\text{SSE}/(n-p)} \]

This is ANOVA F-test
2. Test whether a single $\beta_k=0$

This can be a t-test or a F-test
3. Test whether some of the $\beta_k$'s are 0

$H_0$: $\beta_q = \beta_{q+1} = \ldots = \beta_{p-1} = 0$

Full model:

$Y = \beta_0 + x_1\beta_1 + x_2\beta_2 + \ldots + x_{p-1}\beta_{p-1}$

$SSE(F) = SSE(X_1, \ldots, X_{p-1}) = SSE$

To calculate Degrees of freedom

# of observations = $n$

# of parameters to estimate = $p$ ($\beta_0, \ldots, \beta_{p-1}$)

Hence, $df = n-p$

Reduced Model:

$Y = \beta_0 + x_1\beta_1 + x_2\beta_2 + \ldots + x_{p-1}\beta_{q-1}$

$SSE(R) = SSE(X_1, \ldots, X_{q-1})$

To calculate Degrees of freedom

# of observations = $n$

# of parameters to estimate = $q$ ($\beta_0, \ldots, \beta_{q-1}$)

Hence $df = n-q$

$F = \frac{[SSE(R) - SSE(F) / (n-q-(n-p))] / SSE/(n-p)}{[SSR(X_q, \ldots, X_{p-1} | X_1, \ldots, X_{q-1})/(p-q)] / SSE/(n-p)}$
Example:
Consider a data set, where we are interested in predicting HDL cholesterol based on Weight, Systolic Blood Pressure, Blood Sugar and Triglycerides

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \epsilon$$

<table>
<thead>
<tr>
<th>HDL</th>
<th>$Y_i$</th>
<th>WT</th>
<th>$X_{1i}$</th>
<th>SYS</th>
<th>$X_{2i}$</th>
<th>GLU</th>
<th>$X_{3i}$</th>
<th>TRI</th>
<th>$X_{4i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>53</td>
<td>81</td>
<td>112</td>
<td>132</td>
<td>91</td>
<td>71</td>
<td>82</td>
<td>153</td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>85.7</td>
<td>118</td>
<td>91</td>
<td>82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>91.5</td>
<td>110</td>
<td>87</td>
<td>138</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>113.9</td>
<td>140</td>
<td>100</td>
<td>60</td>
<td>99</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>78.1</td>
<td>130</td>
<td>79</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>91</td>
<td>112</td>
<td>90</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>87.9</td>
<td>118</td>
<td>85</td>
<td>73</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>102.1</td>
<td>128</td>
<td>92</td>
<td>88</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>100.6</td>
<td>120</td>
<td>94</td>
<td>623</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>86.1</td>
<td>130</td>
<td>83</td>
<td>202</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>78.3</td>
<td>118</td>
<td>66</td>
<td>194</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>82.6</td>
<td>114</td>
<td>77</td>
<td>99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>111.2</td>
<td>170</td>
<td>100</td>
<td>206</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>100.9</td>
<td>134</td>
<td>88</td>
<td>93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>86.9</td>
<td>124</td>
<td>82</td>
<td>67</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>80.3</td>
<td>150</td>
<td>81</td>
<td>58</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>90.1</td>
<td>122</td>
<td>96</td>
<td>233</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>99.6</td>
<td>122</td>
<td>83</td>
<td>72</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>81.1</td>
<td>124</td>
<td>101</td>
<td>82</td>
<td>233</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Consider testing the following in terms of Extra Sums of Squares:

Using general Linear Testing: H0: $\beta_4 = 0$ (say) for $X_1, X_2, X_3, X_4$ in the model

Full Model:
\[
Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_4 x_4 + \varepsilon
\]

Full Model:
\[
Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_3 x_3 + \varepsilon
\]

$\text{SSE(Full)} = \text{SSE}(X_1, X_2, X_3, X_4)$
\[
\text{DF}= n-5
\]

Reduced Model
$\text{SSE(Reduced)} = \text{SSE}(X_1, X_2, X_3)$
\[
\text{DF} = n-4
\]

\[
F \text{ test} = \frac{\text{SSE(R)} - \text{SSE(F)}}{[(n-4)-(n-5)]}/\text{SSE(F)}/(n-5)
= \frac{\text{SSR}(X_4| X_1, X_2, X_3)}{1/\text{MSE}}
\]
4. Test whether all $\beta_k = 0$
i.e. $H_0: \beta_1 = \beta_2 = \ldots = \beta_4 = 0$

Full Model:
$Y = \beta_0 + X_1\beta_1 + X_2\beta_2 + \ldots + X_4\beta_4$
$SSE(\text{Full}) = SSE(X_1, \ldots, X_4) = SSE$
$Df = n - 5$

Reduced Model:
$Y = \beta_0$
$Df = n - 1$
$SSE(\text{Reduced}) = SSTO$

$F = \frac{\text{SSTO} - SSE}{(n - 1 - (n - 5))}/\text{SSE}/(n - 5) = \frac{SSR}{4}/\text{SSE}/(n - 5)$

This is ANOVA F-test
5. Test whether some of the $\beta_k$’s are 0

$H_0: \beta_2=\beta_{q+1}=\ldots=\beta_4 =0$

Full model:

$Y = \beta_0 + X_1\beta_1 + X_2\beta_2 + \ldots + X_4\beta_4$

$SSE(F) = SSE(X_1, \ldots, X_4) = SSE$

To calculate Degrees of freedom

# of observations = n
# of parameters to estimate =p ($\beta_0, \ldots, \beta_{p-1}$)

Hence, df = n-5

Reduced Model:

$Y = \beta_0 + x_1\beta_1 + x_2\beta_2$

$SSE(R) = SSE(X_1, \ldots, X_2)$

To calculate Degrees of freedom

# of observations = n
# of parameters to estimate =3 ($\beta_0, \ldots, \beta_4$)

Hence df = n-3

$F= \frac{[SSE(R) - SSE(F)/(n-3-(n-5))] / SSE/(n-5)}{[SSE(X_3, X_4| X_1, X_2)/(2))] / SSE/(n-5)}$
Coefficient of Partial determination:

Measures the marginal contribution of one X variable when all are already in the model.

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

Adjusted R2

\[ R^2a = 1 - \left[ \frac{(SSE)/(n-p)}{(SSTO)/(n-1)} \right] \]

It takes into account the number of predictors in the model and penalizes for too many predictors.
Coefficients of Partial Determination and Correlation

Another use of the Extra Sums of Squares:

\[ R_{Y|123}^2 = \frac{SSE(X1) - SSE(X1X2X3)}{SSE(X2X3)} = \frac{SSR(X1|X2X3)}{SSE(X2X3)} \]

Measures the variation in Y when X2 and X3 were added into the model containing just X1.

In general:

\[ R_{Yab|cd}^2 = \frac{SSR(XaXb|XcXd)}{SSE(XcXd)} \]

To find the correlation just take the square-root of the coefficient of Partial Determination.
Use of the Standardized Multiple Regression Model

Used to control round-off errors and allow calculation of the estimates on the same units.

1. The \((X'X)^{-1}\) matrix is quite subject to round-off errors when the different X’s are on widely different units and orders of magnitude.

2. The slopes are also strictly not comparable to each other if the explanatory variables are in different units.

So often one can “standardize” their Y and X variables by

\[
\frac{Y_i - \bar{Y}}{s_Y} = Y^*, \quad \frac{X_{ik} - \bar{X}}{s_{xk}} = X_{k}^*
\]

Now regressing \(Y^*\) to the \(X_k^*\) allow the comparison of the slopes directly.

The model:

\[
Y_i^* = \beta_1^* X_1^* + \ldots + \beta_{p-1}^* X_{p-1}^*
\]
By definition there is no-intercept in this model. However the original intercept CAN be found out to be:

\[ \beta_0 = \bar{Y} - \beta_1 \bar{X}_1 - \ldots - \beta_{p-1} \bar{X}_{p-1} \]

And we can get back the original slopes using the relationship:

\[ \beta_k^* = \frac{s_y}{s_{xk}} \beta_k \]

Here the \( X'X \) matrix is really the \( r_{xx} \) matrix or the correlation matrix among the \( X \)'s.

And the \( X'Y \) vector is the \( r_{yx} \) vector (vector of simple correlations between the \( Y \) and the individual \( X \)'s).