Main Application of Multiple Regression

- Isolating the effect of a variable by adjusting for the effect of other variables
- Regressing health index versus wealth gives a negative slope for wealth effect
  - The higher the wealth the lower the health
  - What one does when one fits a simple linear regression model
- Regressing health index versus wealth while adjusting for age gives a positive slope for wealth
  - The higher the wealth the higher the health when one adjusts for the effect of the confounder age
- Multiple linear regression models allow us to determine the effect of regressors while adjusting
  - For confounding
  - For effect modification
Why Do Regression Coefficients Have the Wrong Sign?

\[
\hat{y} = 1.835 + 0.463x_1 \\
\hat{y} = 1.036 - 1.222x_1 + 3.649x_2
\]

<table>
<thead>
<tr>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$y$</th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>8</td>
<td>7</td>
<td>4</td>
<td>6</td>
</tr>
</tbody>
</table>

Multiple Linear Regression

Model: \( y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \epsilon \)

Model is called “linear” because it is linear in the \( \beta \)'s not necessarily in the \( x \)'s. So

\[
y = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_k x^k + \epsilon
\]

is considered a linear model and so is

\[
y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 x_1^2 + \beta_5 x_2^2 + \epsilon
\]
A Probabilistic Model for Multiple Linear Regression

\[ Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + \varepsilon_i, \quad i = 1, 2, \ldots, n \]

where \( \varepsilon_i \) is a random error with \( E(\varepsilon_i) = 0 \), and \( \beta_0, \beta_1, \ldots, \beta_k \) are unknown parameters. Usually we assume that the \( \varepsilon_i \) are independent \( N(0, \sigma^2) \) r.v.'s.

This implies that \( Y_i \)'s are independent \( N(\mu_i, \sigma^2) \) r.v.'s with

\[ \mu_i = E(Y_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} \]

Tire Wear Example

\[ k = 2 \quad \rightarrow \quad x_{i1} \quad x_{i2} \quad y_i \]

\[ n = 9 \]

<table>
<thead>
<tr>
<th></th>
<th>Mileage (in 1000 miles)</th>
<th>Mileage^2</th>
<th>Groove Depth (in mil)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>354.33</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>16</td>
<td>329.5</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>64</td>
<td>291</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>144</td>
<td>255.17</td>
</tr>
<tr>
<td>5</td>
<td>14</td>
<td>196</td>
<td>220.30</td>
</tr>
<tr>
<td>6</td>
<td>18</td>
<td>324</td>
<td>204.83</td>
</tr>
<tr>
<td>7</td>
<td>20</td>
<td>400</td>
<td>170</td>
</tr>
<tr>
<td>8</td>
<td>22</td>
<td>484</td>
<td>165.83</td>
</tr>
<tr>
<td>9</td>
<td>24</td>
<td>576</td>
<td>150.30</td>
</tr>
</tbody>
</table>

\[ i = 1, 2, \ldots, n = 9 \]
Least Squares Fit

Minimize:

\[ Q = \sum_{i=1}^{n} \left[ y_i - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik}) \right]^2 \]

By solving the following equations:

\[ \frac{\partial Q}{\partial \beta_0} = 0, \quad \frac{\partial Q}{\partial \beta_1} = 0, \ldots, \quad \frac{\partial Q}{\partial \beta_k} = 0 \]

Solution are Least Square regression estimates:

\[ \hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k \]

Sums of Squares

\[ \hat{y}_i = \hat{\beta}_0 + \hat{\beta}_0 x_{i1} + \ldots + \hat{\beta}_k x_{ik} \quad i = 1, 2, \ldots, n \]

Error Sum of Squares (SSE) = \[ \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \]

Total Sum of Squares (SST) = \[ \sum_{i=1}^{n} (y_i - \bar{y})^2 \]

Regression Sum of Squares (SSR) = \[ \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \]

\[ \sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \]

\[ SST = SSR + SSE \quad \text{As also obtained for simple linear regression} \]
Coefficient of Multiple Determination

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

Percentage of the total variation in Y that is explained by the model

\[ r = \sqrt{r^2}, \text{ is called the} \]

multiple correlation coefficient.

Remark: $R^2$ square always increases with the number of variables included in the model, but this does not mean that more variables always leads to a “better” model.

The Tire Wear Example: Quadratic Fit

Using Fit Y by X platform:

<table>
<thead>
<tr>
<th>Message (in 1000 Miles)</th>
<th>Grove Depth (in mils)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>16</td>
</tr>
<tr>
<td>6</td>
<td>20</td>
</tr>
<tr>
<td>7</td>
<td>24</td>
</tr>
<tr>
<td>8</td>
<td>28</td>
</tr>
<tr>
<td>9</td>
<td>30</td>
</tr>
</tbody>
</table>

Rivariate Fit of Grove Depth (in mils)

- Show Points
- Fit Mean
- Fit Line
- Fit Polynomial
- Fit Special...

Fit Polynomial -- 2.quadratic
Fit Special... 3.cubic
Tire Wear Example: JMP Output

Notice that the polynomial is centered at the average of Mileage = 16.

We used the Fit Y by X platform of JMP to reduce multicollinearity and differences in order of magnitudes among different powers of x.

Tire Wear Example: JMP

In this method of fitting, using the Fit Model platform, JMP does not know that Mileage^2 is the square of Mileage.

The Mileage^2 column was created using JMP's calculator.
Tire Wear Example: JMP Output

**Summary of Fit**
- RSquare: 0.990000
- Rsquare Adj: 0.994777
- Root Mean Square Error: 5.935535
- Mean of Response: 244.1497
- Observations (or Sum Wgts): 9

**Analysis of Variance**
<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>2</td>
<td>53209.471</td>
<td>26604.7</td>
<td>762.3283</td>
</tr>
<tr>
<td>Error</td>
<td>6</td>
<td>206.293</td>
<td>34.9</td>
<td>Prob &gt; F</td>
</tr>
<tr>
<td>C. Total</td>
<td>8</td>
<td>53415.750</td>
<td></td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

**Parameter Estimates**
- Intercept: 396.26465, Std Error: 4.799865, t Ratio: 80.47, Prob > |<.0001
- Mileage (in 1000 Miles): -12.77238, Std Error: 0.330490, t Ratio: -38.26, Prob > |<.0001
- Mileage^2: 0.1718173, Std Error: 0.210332, t Ratio: 8.16, Prob > |0.0002

Polynomial is not centered here. Why?

---

**JMP Fit Model Platform**

Square term is added by highlighting Mileage in both “Select Columns” and “Construct Model Effects” and then clicking the Cross button.

Polynomials are centered by default.
Fit Model Platform Output

Statistical Inference on Individual $\beta$’s

$$
\frac{\hat{\beta}_j - \beta_j}{SE(\hat{\beta}_j)} \sim T_{n-(k+1)} \ (j = 0,1,\ldots,k)
$$

100(1-\alpha)\% CI for $\beta_j$:

$$
\hat{\beta}_j \pm t_{n-(k+1),\alpha/2}SE(\hat{\beta}_j)
$$

$H_{0j} : \beta_j = 0$ versus $H_{1j} : \beta_j \neq 0$

Reject if $|t_j| = \frac{|\hat{\beta}_j|}{SE(\hat{\beta}_j)} > t_{n-(k+1),\alpha/2}$

Recall that the CI are obtained by right clicking on the parameter estimate area and from the popup menu selecting Columns and then Upper 95% and Lower 95%
Statistical Inference on Simultaneous $\beta$’s

\[ H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0 \text{ vs. } H_1 : \text{At least one } \beta_j \neq 0 \]

Reject if \[ F = \frac{MSR}{MSE} > f_{k,n-(k+1),\alpha} \]

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares (SS)</th>
<th>Degrees of Freedom (d.f.)</th>
<th>Mean Square (MS)</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>SSR</td>
<td>$k$</td>
<td>$\frac{SSR}{k}$</td>
<td>$F = \frac{MSR}{MSE}$</td>
</tr>
<tr>
<td>Error</td>
<td>SSE</td>
<td>$n - (k + 1)$</td>
<td>$\frac{SSE}{n-(k+1)}$</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>SST</td>
<td>$n - 1$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**ANOVA Table for Quadratic Fit of Tire Wear Data.**

Other Test of Hypothesis

\[ H_0 : \beta_{k-m+1} = \ldots = \beta_k = 0 \]

Compares

the full model:

\[ Y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + \varepsilon_i \quad (i = 1, 2, \ldots, n) \]

to

the partial model:

\[ Y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_{k-m} x_{i,k-m} + \varepsilon_i \quad (i = 1, 2, \ldots, n) \]

Does adding additional predictor variables to the model lead to a significant reduction in the Error Sums of Squares (SSE)?
Type I and III Sums of Squares

Exercise 11.3:

Does adding a variable to the model last lead to a significant decrease in the error sum of squares?

Part of JMP’s standard output for the Fit Model platform

Type III Sums of Squares

Adding weight to the model - once brain size and height are in the model - does not help predict IQ
Type I or Sequential Sum of Squares

Decrease in the error sums of squares when variables are added one at the time. Here order matters.

Conclusion: Weight does not provide additional information about IQ once brain size and height are in the model.

Sequential SS Add to the Model SS (SSR)

Model Sums of Square = 5572.744

Type III SS do not add up to the Model SS (SSR)
Variable Selection Methods

In practice, a common problem is that there is a large set of candidate predictor variables. One wants to choose a small subset so that the resulting regression model is simple and has good predictive ability

- Stepwise regression
- Best subsets regression
- Optimality criteria

Take “Stat 572: Applied Linear Models” to learn more

**Reading assignment**: “Section 11.8 – A Strategy for Building a Multiple Regression Model”. This strategy will help you with your project.
Stepwise Regression in JMP

Stepwise regression drops “Weight” from the model.

Sequential Sums of Square

Reduced Model Compared with Full Model

R² always increases with number of variables added to the model, but in this case the increase is negligible. This confirms that weight does not add much additional information to the model once brain weight and height are in the model.

Adjusted R² improves when “Weight” is dropped from the model since it has a penalty for number of variables in the model.

Brain size, height, and weight in the models

Only brain size and height in the model
PRESS Statistic

5. PRESS$_p$ Criterion: The PRESS$_p$ criterion$^5$ evaluates the predictive ability of a postulated model by omitting one observation at a time, fitting the model based on the remaining observations and computing the predicted value for the omitted observation. Suppose that we wish to evaluate a $p$-variable model with variables $x_1, x_2, \ldots, x_p$. We fit this model a total of $n$ times by omitting the $i$th observation for $i = 1, 2, \ldots, n$, each time using only the remaining $n-1$ observations. Denote the LS estimates that result when the $i$th observation is omitted by $\hat{\beta}_{(i)p} \cdot \hat{\beta}_{(i)p}$. Then the predicted value for the omitted observation is $\hat{y}_{(i)p} = \hat{\beta}_{(i)p} + \hat{\beta}_{(i)p}x_{i1} + \cdots + \hat{\beta}_{(i)p}x_{ip}$. The total prediction error sum of squares (PRESS) is given by

$$\text{PRESS}_p = \sum_{i=1}^{n} (\hat{y}_{(i)p} - y_i)^2.$$ 

This quantity is evaluated for each contending model, and the one associated with the minimum value of PRESS$_p$ is chosen.

PRESS in JMP

7/9/2003 Unit 11 - Stat 571 - Ramón V. León 27
Analysis Methods that Are Similar to Those in Simple Linear Regression

- Confidence intervals for the **average response** of Y at a particular value of the regressors
- Prediction intervals for a **future value** of Y for a particular value of the regressors
- Plot of residuals
  - against individual predictor variables included or not in the model
  - against predicted value
  - on normal plot
  - in time order (Durbin-Watson test)
  - Standardized residuals (> 2 indicates an outlier)
- Influential observations if
  \[ h_{ij} > \frac{2(k + 1)}{n} \]

Transformations

- For the predictor variables as well as the response to make the model linear
- A model that can be transformed so that it becomes linear in its unknown parameters is called **intrinsically linear** otherwise it is called a **nonlinear model**
- Unless a common transformation of y linearizes its relationship with all x’s, effort should be focused on transforming the x’s, leaving y untransformed, at least initially.
- It is usually assumed that the random error enters additively into the final form of the model
Illustration of Multicollinearity: Cement Data

Table 11.2 gives data\(^2\) on the heat evolved in calories during hardening of cement on a per gram basis (\(y\)) along with the percentages of four ingredients: tricalcium aluminate (\(x_1\)), tricalcium silicate (\(x_2\)), tetracalcium aluminoferrite (\(x_3\)), and dicalcium silicate (\(x_4\)). We will illustrate the multicollinearity present in these data and its consequences on fitting the regression model.

### Table 11.2 Cement Data

<table>
<thead>
<tr>
<th>No.</th>
<th>(x_1)</th>
<th>(x_2)</th>
<th>(x_3)</th>
<th>(x_4)</th>
<th>(y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>26</td>
<td>6</td>
<td>60</td>
<td>78.5</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>29</td>
<td>15</td>
<td>52</td>
<td>74.3</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>56</td>
<td>8</td>
<td>20</td>
<td>104.3</td>
</tr>
<tr>
<td>4</td>
<td>11</td>
<td>31</td>
<td>8</td>
<td>47</td>
<td>87.6</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>52</td>
<td>6</td>
<td>33</td>
<td>95.9</td>
</tr>
<tr>
<td>6</td>
<td>11</td>
<td>55</td>
<td>9</td>
<td>22</td>
<td>109.2</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>71</td>
<td>17</td>
<td>6</td>
<td>102.7</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>31</td>
<td>22</td>
<td>44</td>
<td>72.5</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>54</td>
<td>18</td>
<td>22</td>
<td>93.1</td>
</tr>
<tr>
<td>10</td>
<td>21</td>
<td>47</td>
<td>4</td>
<td>26</td>
<td>115.9</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>40</td>
<td>23</td>
<td>34</td>
<td>83.8</td>
</tr>
<tr>
<td>12</td>
<td>11</td>
<td>66</td>
<td>9</td>
<td>12</td>
<td>113.3</td>
</tr>
<tr>
<td>13</td>
<td>10</td>
<td>68</td>
<td>8</td>
<td>12</td>
<td>109.4</td>
</tr>
</tbody>
</table>

Note that \(x_1\), \(x_2\), \(x_3\), and \(x_4\) add up to approximately 100% for all observations.

Correlation Matrix: Cement Data

Note that the \(x_1\), \(x_2\), \(x_3\), and \(x_4\) add up to approximately 100% for all observations. This approximate linear relationship among the \(x\)'s results in multicollinearity. **Multicollinearity** means that the predictor variable are linearly related.

### Correlations

<table>
<thead>
<tr>
<th></th>
<th>(x_1)</th>
<th>(x_2)</th>
<th>(x_3)</th>
<th>(x_4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(x_1)</td>
<td>1.0000</td>
<td>0.2286</td>
<td>-0.8241</td>
<td>-0.2454</td>
</tr>
<tr>
<td>(x_2)</td>
<td>0.2286</td>
<td>1.0000</td>
<td>-0.1392</td>
<td>-0.9730</td>
</tr>
<tr>
<td>(x_3)</td>
<td>-0.8241</td>
<td>-0.1392</td>
<td>1.0000</td>
<td>0.0295</td>
</tr>
<tr>
<td>(x_4)</td>
<td>-0.2454</td>
<td>-0.9730</td>
<td>0.0295</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
Scatter Plot Matrix of Cement Data

Multicollinearity

Figure 3.13 (a) A data set with multicollinearity. (b) Orthogonal regressors.
Multicollinearity

- Multicollinearity can cause serious numerical and statistical difficulties in fitting the regression model unless “extra” predictor variables are deleted.
  - Example: If income, expenditure, and savings are used as predictor variables one should be deleted since saving = income – expenditure
- Multicollinearity leads to these problems:
  - The estimates $\hat{\beta}$ are subject to numerical errors and are unreliable. This is reflected in large changes in their magnitudes with small changes in data.
  - Most of the estimated coefficients have large standard errors and as a result are statistically nonsignificant even if the overall F-statistic is significant.

Regression Coefficients in the Presence of Multicollinearity: Cement Data

| Parameter Estimates | Term   | Estimate | Std Error | t Ratio | Prob>|t| |
|---------------------|--------|----------|-----------|---------|----------|
| Intercept           | 62.405369 | 70.07096 | 0.89     | 0.3991  |
| X1                  | 1.5511026 | 0.74477 | 2.08     | 0.0708  |
| X2                  | 0.5101676 | 0.723788 | 0.70     | 0.5009  |
| X3                  | 0.1019094 | 0.754709 | 0.14     | 0.8959  |
| X4                  | -0.144061 | 0.709052 | -0.20    | 0.8441  |

<table>
<thead>
<tr>
<th>Analysis of Variance</th>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Ratio</th>
<th>Prob &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model</td>
<td>4</td>
<td>2667.8994</td>
<td>666.975</td>
<td>111.4792</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>8</td>
<td>47.8636</td>
<td>5.983</td>
<td>Prob &gt; F</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C. Total</td>
<td>12</td>
<td>2715.7631</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notice that all the regression coefficients are nonsignificant, although the overall $F = 111.48$ is highly significant.
Measures of Multicollinearity: Correlations and Variance Inflation Factor (VIF)

The simplest measure is the **correlations matrix** between pairs of predictor variables. A more direct measure is the **VIF**.

\[ VIF_j = \text{Diagonal elements of inverse of correlation matrix} \]

We can show that \[ VIF_j = \frac{1}{1-r_j^2}, \quad i = 1, 2, ..., k \]

where \( r_j^2 \) is the coefficient of multiple determination when regressing \( x_j \) on the remaining \( k-1 \) predictor variables.

Generally, \( VIF_j \) values greater than 10, corresponding to \( r_j^2 > .9 \), are regarded as unacceptable.

---

VIFs for Cement Data

<table>
<thead>
<tr>
<th>Parameter Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Term</td>
</tr>
<tr>
<td>------</td>
</tr>
<tr>
<td>Intercept</td>
</tr>
<tr>
<td>X1</td>
</tr>
<tr>
<td>x2</td>
</tr>
<tr>
<td>x3</td>
</tr>
<tr>
<td>X4</td>
</tr>
</tbody>
</table>

Notice that all the VIFs are greater than 10, a strong indication of multicollinearity.
VIF in JMP: How?

Right click on the Parameter Estimates area. Select VIF from the columns submenu that pops up:

Polynomial Regression Remarks

- **Problems**
  - Power of $x$ are highly correlated: collinearity
  - Powers of $x$ can have large differences in order of magnitude

- **Solutions**
  - Center the $x$ around the mean:
    \[ x - \bar{x} \]
  - Standardizing the $x$ is even better:
    \[ \frac{x - \bar{x}}{s_x} \]
Dummy Predictor Variables

Many applications involve categorical predictor variables, e.g. gender, season, prognosis (poor, average, good)

• For ordinal variables may use scores, e.g. use 1, 2, and 3 for the prognosis categories
• For nominal variables with \( c \) categories use \( c-1 \) indicator variables \( x_1, x_2, \ldots, x_{c-1} \) called dummy variables.

Example:

\[
\begin{align*}
x_1 &= \begin{cases} 
1 & \text{if Winter} \\
0 & \text{if another season} 
\end{cases} \\
x_2 &= \begin{cases} 
1 & \text{if Spring} \\
0 & \text{if another season} 
\end{cases} \\
x_3 &= \begin{cases} 
1 & \text{if Summer} \\
0 & \text{if another season} 
\end{cases} \\
x_4 &= \begin{cases} 
1 & \text{if Fall} \\
0 & \text{if another season} 
\end{cases}
\end{align*}
\]

Model:

\[ Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon \]

Fall is taken as reference point: \( \beta_0 \)

Why 1-c Dummy Variables?

If

\[
\begin{align*}
x_1 &= \begin{cases} 
1 & \text{if Winter} \\
0 & \text{if another season} 
\end{cases} \\
x_2 &= \begin{cases} 
1 & \text{if Spring} \\
0 & \text{if another season} 
\end{cases} \\
x_3 &= \begin{cases} 
1 & \text{if Summer} \\
0 & \text{if another season} 
\end{cases} \\
x_4 &= \begin{cases} 
1 & \text{if Fall} \\
0 & \text{if another season} 
\end{cases}
\end{align*}
\]

Then \( x_1 + x_2 + x_3 + x_4 = 1 \)

Multicollinearity
JMP’s Dummy Variables

\[
z_1 = \begin{cases} 
1 & \text{if Winter} \\
-1 & \text{if Fall} \\
0 & \text{if another season}
\end{cases} \\
z_2 = \begin{cases} 
1 & \text{if Spring} \\
-1 & \text{if Fall} \\
0 & \text{if another season}
\end{cases} \\
z_3 = \begin{cases} 
1 & \text{if Summer} \\
-1 & \text{if Fall} \\
0 & \text{if another season}
\end{cases}
\]

\[Y = \beta_0 + \beta_1(z_1) + \beta_2(z_2) + \beta_3(z_3) + \varepsilon\]

Four season average is taken as reference point: \(\beta_0\)

\[
Y_{\text{Fall}} = \beta_0 + \beta_1(-1) + \beta_2(-1) + \beta_3(-1) + \varepsilon \\
+ Y_{\text{Winter}} = \beta_0 + \beta_1(1) + \varepsilon \\
+ Y_{\text{Spring}} = \beta_0 + \beta_2(1) + \varepsilon \\
+ Y_{\text{Summer}} = \beta_0 + \beta_3(1) + \varepsilon
\]

\[\bar{Y} = \beta_0 + \frac{\varepsilon}{4}\]

Quarterly Sales of Soda Cans
(in Millions of Cans)

<table>
<thead>
<tr>
<th>Quarter</th>
<th>Years</th>
<th>Season</th>
<th>Sales</th>
<th>(x_1)</th>
<th>(x_2)</th>
<th>(x_3)</th>
<th>(z_1)</th>
<th>(z_2)</th>
<th>(z_3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>Winter</td>
<td>41.2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>Spring</td>
<td>50.2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>Summer</td>
<td>59.5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>Fall</td>
<td>41.3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>Winter</td>
<td>44.6</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>Spring</td>
<td>59.9</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>Summer</td>
<td>59.3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>Fall</td>
<td>54.7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>Winter</td>
<td>54.7</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>Spring</td>
<td>74.2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>11</td>
<td>Summer</td>
<td>79.4</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>12</td>
<td>Fall</td>
<td>52.7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>13</td>
<td>13</td>
<td>Winter</td>
<td>52.7</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>14</td>
<td>Spring</td>
<td>79.0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>15</td>
<td>Summer</td>
<td>82.6</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>16</td>
<td>Fall</td>
<td>85.6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
</tbody>
</table>

* The \(x\)’s are the textbook (SAS) dummy variables.  
* The \(z\)’s are JMP’s dummy variables.
Runs Chart of Soda Can Sales

Try Model:

\[ Y = \beta_0 + \beta_1 QUARTER + \beta_2 x_1 + \beta_3 x_2 + \beta_4 x_3 + \varepsilon \]

Linear time trend and cyclical seasonal trend

JMP Analysis: Textbook (SAS) Dummy Variables

After adjusting for the linear time trend, the Spring and Summer sales differ significantly from the Fall sales, but those of Winter do not.

Notice that these variables are declared continuous in JMP.
After adjusting for the linear time trend, Winter, Spring, and Summer (and Fall) differ significantly from the average of the four quarters.

Nominal

To make “Fall” the “-1” z variable.
How to Tell Season Effect is Important*

<table>
<thead>
<tr>
<th>Source</th>
<th>Nparm</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>F Ratio</th>
<th>Prob &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quarter</td>
<td>1</td>
<td>1</td>
<td>1602.0500</td>
<td>182.0605</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Season</td>
<td>3</td>
<td>3</td>
<td>998.9424</td>
<td>37.8407</td>
<td>&lt;.0001*</td>
</tr>
</tbody>
</table>

There is statistical evidence of a seasonal effect.

*Part the output of the command in the previous visual

Confidence and Prediction Intervals

Notice that one should not use the “Mean Confidence Interval”

<table>
<thead>
<tr>
<th>Quarter</th>
<th>Years</th>
<th>Season</th>
<th>Sales</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>15</td>
<td>Fall</td>
<td>65.6</td>
</tr>
<tr>
<td>17</td>
<td>17</td>
<td>Winter</td>
<td>4</td>
</tr>
</tbody>
</table>

Predicted sales for seventeenth quarter (next Winter) in last line
Time Series Reference


One Dummy and One Continuous Predictor Variables

\[ E(\text{Weight}) = \beta_0 + \beta_1 \text{SEX} + \beta_2 \text{HEIGHT} \]

where \( \text{SEX} = 0 \) for females and 1 for males.

Equivalent formulation:

\[ E(\text{Weight}) = \begin{cases} 
\beta_0 + \beta_2 \text{HEIGHT} & \text{for females} \\
(\beta_0 + \beta_1) + \beta_2 \text{HEIGHT} & \text{for males}
\end{cases} \]

Lines have the same slope

Height is an effect confounder for the gender effect.

No interaction term in the model
One Dummy and One Continuous Predictor Variables

\[ E(\text{Weight}) = \beta_0 + \beta_1 \text{SEX} + \beta_2 \text{HEIGHT} \]

where \( \text{SEX} = 0 \) for females and 1 for males.

Equivalent formulation:

\[ E(\text{Weight}) = \begin{cases} 
\beta_0 + \beta_2 \text{HEIGHT} & \text{for females} \\
(\beta_0 + \beta_1) + \beta_2 \text{HEIGHT} & \text{for males}
\end{cases} \]

No interaction term in the model

Lines have the same slope

Height is an effect confounder for the gender effect.

Interaction

\[ E(\text{Weight}) = \beta_0 + \beta_1 \text{SEX} + \beta_2 \text{HEIGHT} + \beta_3 \text{SEX} \times \text{HEIGHT} \]

where \( \text{SEX} = 0 \) for females and 1 for males.

Equivalent formulation:

\[ E(\text{Weight}) = \begin{cases} 
\beta_0 + \beta_2 \text{HEIGHT} & \text{for females} \\
(\beta_0 + \beta_1) + (\beta_2 + \beta_3) \text{HEIGHT} & \text{for males}
\end{cases} \]

Interaction Term

Lines have different slopes

Height is an effect modifier for the gender effect

Interactions are also important to consider among continuous and among categorical variables. This is a key step in model building.
Considering Interactions

- **Fit Model**
  - **Model Specification**
    - **Select Columns**
      - P-kg
      - Brain Size
      - Height
      - Weight
    - **Add**
    - **Construct Model Effects**
      - Brain Size
      - Brain Size*Height
      - Brain Size*Weight
      - Height
      - Weight

- **Stepwise Fit**
  - **Stepwise Regression Control**
    - **Prob to Enter** 0.250
    - **Prob to Leave** 0.100
    - **Direction** Forward
    - **Rules** Continue
  - **Current Estimates**
    | Parameter       | Estimate | DF | SS | F Ratio | P Value |
    |-----------------|----------|----|----|---------|---------|
    | Intercept       | 111.525623 | 1 | 0 | 0.0000 | 0.0000  |
    | Brain Size      | 2.009454  | 2 | 549.049 | 9.903   | 0.0000  |
    | Height          | -2.725961 | 2 | 2075.821 | 3.670   | 0.0306  |
    | Brain Size*Height| -0.802386 | 1 | 0.27305 | 0.001   | 0.9761  |
    | Weight          | 0        | 1 | 0.007438 | 0.000     | 0.9965  |
    | Brain Size*Weight| 0.014362 | 2 | 9.344322 | 0.005   | 0.9983  |
    | Height*Weight   | 0        | 2 | 271.0262 | 0.002   | 0.9767  |

- **Stop History**
<table>
<thead>
<tr>
<th>Step</th>
<th>Parameter</th>
<th>Action</th>
<th>&quot;Sig.Prob&quot;</th>
<th>Sex SS</th>
<th>R Square</th>
<th>Cp</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(Brain Size*Height)</td>
<td>Entered</td>
<td>0.0082</td>
<td>5570.315</td>
<td>0.2950</td>
<td>1.0915</td>
<td>4</td>
</tr>
</tbody>
</table>
The interaction of “Brain Size” and “Height” is not significant
Fallacy of Doing Simple Linear Regression When Multiple Linear Regression is Called For

- Regressing health index versus wealth gives a negative slope for wealth effect
  - The higher the wealth the lower the health
  - This is what one does when one fits a simple linear regression model
- Regressing health index versus wealth while adjusting for age gives a positive slope for wealth
  - The higher the wealth the higher the health
  - This is what one does when one fits a multiple linear regression model of the health index versus wealth and age

Logistic Regression and Logit Models

Logistic Regression used when the response variable is binary (or more generally categorical), e.g., a patient survives or dies.

Model:
\[
\ln \left( \frac{P(Y = 1 \mid x_1, x_2, \ldots, x_k)}{P(Y = 0 \mid x_1, x_2, \ldots, x_k)} \right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k
\]

Equivalently,
\[
P(Y = 1 \mid x_1, x_2, \ldots, x_k) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k}}
\]
### Age and Coronary Heart Disease: Status of 100 Subjects

<table>
<thead>
<tr>
<th>Age</th>
<th>CHD</th>
<th>Age</th>
<th>CHD</th>
<th>Age</th>
<th>CHD</th>
<th>Age</th>
<th>CHD</th>
<th>Age</th>
<th>CHD</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0</td>
<td>34</td>
<td>0</td>
<td>41</td>
<td>0</td>
<td>48</td>
<td>1</td>
<td>57</td>
<td>0</td>
</tr>
<tr>
<td>23</td>
<td>0</td>
<td>34</td>
<td>0</td>
<td>42</td>
<td>0</td>
<td>48</td>
<td>1</td>
<td>57</td>
<td>1</td>
</tr>
<tr>
<td>24</td>
<td>0</td>
<td>34</td>
<td>1</td>
<td>42</td>
<td>0</td>
<td>49</td>
<td>0</td>
<td>57</td>
<td>1</td>
</tr>
<tr>
<td>25</td>
<td>0</td>
<td>34</td>
<td>0</td>
<td>42</td>
<td>0</td>
<td>49</td>
<td>0</td>
<td>57</td>
<td>1</td>
</tr>
<tr>
<td>25</td>
<td>1</td>
<td>34</td>
<td>0</td>
<td>42</td>
<td>1</td>
<td>49</td>
<td>1</td>
<td>57</td>
<td>1</td>
</tr>
<tr>
<td>26</td>
<td>0</td>
<td>35</td>
<td>0</td>
<td>43</td>
<td>0</td>
<td>50</td>
<td>0</td>
<td>58</td>
<td>0</td>
</tr>
<tr>
<td>26</td>
<td>0</td>
<td>35</td>
<td>0</td>
<td>43</td>
<td>0</td>
<td>50</td>
<td>1</td>
<td>58</td>
<td>1</td>
</tr>
<tr>
<td>28</td>
<td>0</td>
<td>36</td>
<td>0</td>
<td>43</td>
<td>1</td>
<td>51</td>
<td>0</td>
<td>58</td>
<td>1</td>
</tr>
<tr>
<td>28</td>
<td>0</td>
<td>36</td>
<td>1</td>
<td>44</td>
<td>0</td>
<td>52</td>
<td>0</td>
<td>59</td>
<td>1</td>
</tr>
<tr>
<td>29</td>
<td>0</td>
<td>36</td>
<td>0</td>
<td>44</td>
<td>0</td>
<td>52</td>
<td>1</td>
<td>59</td>
<td>1</td>
</tr>
<tr>
<td>30</td>
<td>0</td>
<td>37</td>
<td>0</td>
<td>44</td>
<td>1</td>
<td>53</td>
<td>1</td>
<td>60</td>
<td>0</td>
</tr>
<tr>
<td>30</td>
<td>0</td>
<td>37</td>
<td>1</td>
<td>44</td>
<td>1</td>
<td>53</td>
<td>1</td>
<td>60</td>
<td>1</td>
</tr>
<tr>
<td>30</td>
<td>0</td>
<td>37</td>
<td>0</td>
<td>45</td>
<td>0</td>
<td>54</td>
<td>1</td>
<td>61</td>
<td>1</td>
</tr>
<tr>
<td>30</td>
<td>0</td>
<td>38</td>
<td>0</td>
<td>45</td>
<td>1</td>
<td>55</td>
<td>0</td>
<td>62</td>
<td>1</td>
</tr>
<tr>
<td>30</td>
<td>0</td>
<td>38</td>
<td>0</td>
<td>46</td>
<td>0</td>
<td>55</td>
<td>1</td>
<td>62</td>
<td>1</td>
</tr>
<tr>
<td>30</td>
<td>1</td>
<td>39</td>
<td>0</td>
<td>46</td>
<td>1</td>
<td>55</td>
<td>1</td>
<td>63</td>
<td>1</td>
</tr>
<tr>
<td>32</td>
<td>0</td>
<td>39</td>
<td>1</td>
<td>47</td>
<td>0</td>
<td>56</td>
<td>1</td>
<td>64</td>
<td>0</td>
</tr>
<tr>
<td>32</td>
<td>0</td>
<td>40</td>
<td>0</td>
<td>47</td>
<td>0</td>
<td>56</td>
<td>1</td>
<td>64</td>
<td>1</td>
</tr>
<tr>
<td>33</td>
<td>0</td>
<td>40</td>
<td>1</td>
<td>47</td>
<td>1</td>
<td>56</td>
<td>1</td>
<td>65</td>
<td>1</td>
</tr>
<tr>
<td>33</td>
<td>0</td>
<td>41</td>
<td>0</td>
<td>48</td>
<td>0</td>
<td>57</td>
<td>0</td>
<td>69</td>
<td>1</td>
</tr>
</tbody>
</table>

### Coronary Heart Disease by Age

![Coronary Heart Disease by Age](image_url)
Logistic Regression in JMP

Notice that CHD is declared nominal.

JMP Output

The probability of coronary heart disease increases with age.

Ask JMP with the question mark tool for the complete interpretation of this plot.
JMP Output

**Whole Model Test**

<table>
<thead>
<tr>
<th>Model</th>
<th>-LogLikelihood</th>
<th>DF</th>
<th>ChiSquare</th>
<th>Prob&gt;ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difference</td>
<td>14.654945</td>
<td>1</td>
<td>29.30989</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Full</td>
<td>53.676546</td>
<td>29.30989</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>Reduced</td>
<td>68.331491</td>
<td>29.30989</td>
<td>&lt;.0001</td>
<td></td>
</tr>
</tbody>
</table>

RSquare (U) 0.2145
Observations (or Sum Wgts) 100
Converged by Gradient

**Parameter Estimates**

<table>
<thead>
<tr>
<th>Term</th>
<th>Estimate</th>
<th>Std Error</th>
<th>ChiSquare</th>
<th>Prob&gt;ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>5.30945302</td>
<td>1.1336546</td>
<td>21.94</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>AGE</td>
<td>-0.1109211</td>
<td>0.0240598</td>
<td>21.25</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

For log odds of 0/1

Counterpart to the Analysis of Variance table in multiple linear regression with similar interpretation.

Logistic Regression Using JMP’s Fit Model Platform:
How to do Multiple Logistic Regression
### Do You Need to Know More?


**Recommended References:**

- **Introduction to Linear Regression Analysis, Third Edition**
  A very well-written introduction to regression including some modern methods.

- **Applied Logistic Regression, Second Edition**
  by David W. Hosmer and Stanley Lemeshow. Wiley Interscience.
  A great introduction to logistic regression for those already familiar with the elements of multiple regression analysis.