Comp 260: Computational Models and Methods

Linear Regression: $f(x)$ minimizes $\sum_{i=1}^{n} \left( y_i - f(x_i) \right)^2$

Error sum of squares (SSE): $\sum_{i=1}^{n} \left( y_i - f(x_i) \right)^2$

Total corrected sum of squares (SST): $\sum_{i=1}^{n} \left( y_i - \bar{y} \right)^2$

Regression sum of squares: $SSR = SST - SSE$

$SSR = SST - SSE$ reflects the amount of variation in the $y$ values explained by the linear regression line $y = a \cdot x + b$ when compared with the variation of the $y$ values about the line $\bar{y}$.

$SST \geq SSE$

Coefficient of regression $R^2 = 1 - \frac{SSE}{SST}$

Reasonableness of fit can also be determined by examining a plot of the residuals $y_i - f(x_i)$.

Ponderosa Pine Example (page 235)

Two Possible Modeling Scenarios
- height is proportional to diameter: $V \propto d^3$
- height is independent of diameter: $V \propto d^2$

Four (Linear Regression) Models (use Fit[])

<table>
<thead>
<tr>
<th>Model</th>
<th>SSE</th>
<th>SSR</th>
<th>SST</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V = 0.00432d^3$</td>
<td>3741.85</td>
<td>164862</td>
<td>168604</td>
<td>0.977807</td>
</tr>
<tr>
<td>$V = 0.00426d^3 + 2.08$</td>
<td>3711.81</td>
<td>164892</td>
<td>168904</td>
<td>0.977985</td>
</tr>
<tr>
<td>$V = 0.152d^2$</td>
<td>12894.6</td>
<td>155709</td>
<td>168904</td>
<td>0.923522</td>
</tr>
<tr>
<td>$V = 0.194d^2 - 45.7$</td>
<td>3909.82</td>
<td>164694</td>
<td>168904</td>
<td>0.970811</td>
</tr>
</tbody>
</table>

Notes