Sums of Squares

- Sums of squares are used to quantify variation in the response variable.
- In general, the larger a sum of squares is the more variation there is.

Variation in Population

- There is a lot of variation in the population values.
- This can be quantified by:
  - the sample standard deviation, \( s = 96.2357 \)
  - the sample variance, \( s^2 = 9261.31 \)
The sample variance

The formula for the sample variance is:

\[ s^2 = \frac{\sum(Y - \bar{Y})^2}{n - 1} = \frac{SS_{C.Total}}{df_{C.Total}} \]

Sum of Squares: C. Total

- \( SS_{C.Total} \) takes each of the deviations from the sample mean, squares them and adds them together.
- This is part of how one quantifies the total variation in the response.
Simple Linear Regression

- Simple linear regression uses an explanatory variable to help explain variation in the response.
- Predictions using the simple linear regression should be closer, on average, to the observed values of the response.

SLR of Population on Year

Predicted Population = \(-2480.85 + 1.360\times\text{Year}\)

For 2010.
- Population = 308.746 million
- Predicted Population = 252.750 million
- Residual = 308.746 - 252.750 = 55.996 million
Stat 301 – Sums of Squares

C. Total = Model + Residual

- The deviation from the mean is split into two pieces.
  - A piece explained by the model.
  - A piece that is not explained, residual error.

\[ (Y - \bar{Y}) = (\bar{Y} - \bar{Y}) + (Y - \bar{Y}) \]

\[ \text{SS}_C. \text{Total} = \text{SS}_\text{Model} + \text{SS}_\text{Error} \]

\[ \sum (Y - \bar{Y})^2 = \sum (\bar{Y} - \bar{Y})^2 + \sum (Y - \bar{Y})^2 \]

Rsquare (R^2)

\[ R^2 = \frac{\text{SS}_\text{Model}}{\text{SS}_C.\text{Total}} = \frac{187300.39}{203748.80} = 0.919 \]

- 91.9% of the variation in population is explained by the simple linear regression model with year.
Quadratic Model

- Does adding a Year² term to Year explain more of the variation in population?
- Is this additional explained variation statistically significant?

Change in RSquare

- Quadratic Model: Year, Year²
  - RSquare = 0.999099
- Linear Model: Year
  - RSquare = 0.919271
- Change in RSquare = 0.079828
Stat 301 – Sums of Squares

### Change in RSquare

\[
0.079828 = \frac{SS_{Year^2|Year}}{SS_{C.Total}} = \frac{SS_{Year^2|Year}}{203748.80}
\]

- Adding \( Year^2 \) to \( Year \) increases the \( SS_{Model} \) by 16264.86

### Partial F-test

- Adding \( Year^2 \) uses 1 df.
- \[
F = \frac{MS_{Year^2|Year}}{MS_{Error}} = \frac{(16264.86/1)}{9.177}
\]
- \( F = 1772.4 \)
- P-value < 0.0001
- \( Year^2 \) is a statistically significant addition to the model.

### Summary of Fit

- RSquare: 0.999009
- RSquare Adj: 0.999099
- Root Mean Square Error: 3.029323
- Mean of Response: 103.9827
- Observations (or Sum Wgts): 23

### Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Mean Square</th>
<th>F Ratio</th>
<th>Prob &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>2</td>
<td>101783</td>
<td>1772.391</td>
<td>&lt;0.0001*</td>
</tr>
<tr>
<td>Error</td>
<td>20</td>
<td>9.17758</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. Total</td>
<td>22</td>
<td>203748.80</td>
<td>&lt;0.0001*</td>
<td></td>
</tr>
</tbody>
</table>

### Effect Tests

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Mean Square</th>
<th>F Ratio</th>
<th>Prob &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
<td>1</td>
<td>16234.47</td>
<td>1772.391</td>
<td>&lt;0.0001*</td>
</tr>
<tr>
<td>Year*Year</td>
<td>1</td>
<td>16234.47</td>
<td>1772.391</td>
<td>&lt;0.0001*</td>
</tr>
</tbody>
</table>