\[ b = \frac{S_{xy}}{S_{xx}} \]

\[ SSE = S_{yy} - 2bS_{xy} + \frac{(S_{xy})^2}{S_{xx}} \cdot S_{xx} \]

\[ = S_{yy} - 2bS_{xy} + bS_{xy} \]

\[ SSE = S_{yy} - bS_{xy} \]

The estimate of the variance of the points about the line is:

\[ s^2_{y|x} = \frac{SSE}{n-2} = \frac{S_{yy} - bS_{xy}}{n-2} \]

This quantity is a measure of the scatter of experimental points around the line.
Example 2: For the data of example 0, calculate the standard deviation of points about the regression line, then plot residuals against x.

\[ \hat{y} = a + bx \]
\[ \hat{y} = 1.0511 + 1.36456x \] (from ex. 0)

Residual \( e_i = y_i - \hat{y} \)

\[ s_{y|x} = \sqrt{\frac{SSE}{n-2}} \]

<table>
<thead>
<tr>
<th>( x_i )</th>
<th>( y_i )</th>
<th>( \hat{y} )</th>
<th>( e_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>3.85</td>
<td>1.05</td>
<td>+2.8</td>
</tr>
<tr>
<td>1</td>
<td>0.03</td>
<td>2.41</td>
<td>-2.38</td>
</tr>
<tr>
<td>2</td>
<td>3.55</td>
<td>3.77</td>
<td>-0.27</td>
</tr>
<tr>
<td>3</td>
<td>6.13</td>
<td>5.13</td>
<td>+1.0</td>
</tr>
<tr>
<td>4</td>
<td>4.07</td>
<td>6.49</td>
<td>-2.42</td>
</tr>
<tr>
<td>5</td>
<td>7.67</td>
<td>7.85</td>
<td>-0.78</td>
</tr>
<tr>
<td>6</td>
<td>8.66</td>
<td>9.21</td>
<td>-0.55</td>
</tr>
<tr>
<td>7</td>
<td>11.65</td>
<td>10.57</td>
<td>+1.08</td>
</tr>
<tr>
<td>8</td>
<td>15.23</td>
<td>11.93</td>
<td>+3.3</td>
</tr>
<tr>
<td>9</td>
<td>12.29</td>
<td>13.56</td>
<td>-1.27</td>
</tr>
<tr>
<td>10</td>
<td>14.74</td>
<td>14.65</td>
<td>+0.09</td>
</tr>
<tr>
<td>11</td>
<td>16.02</td>
<td>16.01</td>
<td>+0.01</td>
</tr>
<tr>
<td>12</td>
<td>17.02</td>
<td>17.22</td>
<td>-0.51</td>
</tr>
</tbody>
</table>
\[ \text{SSE} = S_{yy} - b \cdot S_{xy} \]
\[ = 373.5436 - 1.36456 \times (248.35) \]
\[ = 34.655 \]

\[ S_{y|x} = \sqrt{\frac{\text{SSE}}{n - 2}} = \sqrt{\frac{34.655}{12 - 2}} \]

\[ S_{y|x} = 1.86 \]

[Scatter plot with labeled axes]
Relation forms

1. Straight line through origin \( \rightarrow y = mx \)

2. Other single constant forms are all transformable to straight line through origin:
   \[ y = m e^x \quad \text{Define } Y = y, \quad X = e^x \quad \rightarrow \quad Y = m X. \]

3. Straight line \( \rightarrow y = a_0 + a_1 x \)

4. Straight line forms:
   - Two-constant relations may be transformed to straight line:
     \[ y = a e^{bx} \quad \text{exponential} \quad \rightarrow \quad \ln y = \ln a + bx \]
     \[ y = a x^b \quad \text{Power} \quad \rightarrow \quad \ln y = \ln a + b \ln x \]
     \[ y = \frac{1}{a_0 + a_1 x} \quad \text{hyperbola} \quad \rightarrow \quad \frac{1}{y} = a_0 + a_1 x. \]

5. Higher constant relations:
   These may be polynomials or other forms that contain more than two constants. It is not usually possible to transform them into straight line form.
e.g. \( y = a_0 + a_1 x + a_2 x^2 \) 2nd degree Polynomial

\[ y = a + be^x \] modified exponential

\[ \frac{a}{y} = b + Cx \rightarrow \frac{1}{y} = \frac{b}{a} + \frac{C}{a} x \text{ (st. line form)} \]

Ex.

Transform \( p = \exp \left( a + \frac{1}{bx} \right) \) and define parameters:

\[ \ln p = a + \frac{1}{bx} \quad \therefore \quad y = \ln p, \quad x = \frac{1}{x} \]

\[ A_0 = a, \quad A_1 = \frac{1}{b} \]
Examples

1) Fit the following data to a straight line

<table>
<thead>
<tr>
<th>time</th>
<th>0</th>
<th>3</th>
<th>5</th>
<th>8</th>
<th>10</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>speed</td>
<td>0.28</td>
<td>11.2</td>
<td>18.3</td>
<td>29.1</td>
<td>36.2</td>
<td>43.4</td>
</tr>
</tbody>
</table>

Sln.

\[ N = 6 \quad \Sigma x_i = 38 \quad \Sigma y_i = 138.48 \]
\[ \Sigma x_i^2 = 342 \quad \Sigma y_i^2 = 4501.2 \quad \Sigma x_i y_i = 1240.7 \]
\[ \bar{y} = 23.08, \quad \bar{x} = 6.33 \]

\[ b = \frac{\Sigma xy}{\Sigma xx} \quad \quad a = \bar{y} - b\bar{x} \]

\[ \Sigma xy = \Sigma xy - \frac{1}{n} (\Sigma x)(\Sigma y) \]
\[ = 1240.7 - \frac{1}{6} (38)(138.48) = 363.7 \]

\[ \Sigma xx = \Sigma x_i^2 - \frac{1}{n} (\Sigma x)^2 = 342 - \frac{1}{6} (38)^2 = 101.3 \]

\[ b = \frac{363.7}{101.3} = 3.59, \quad a = 23.08 - 3.59 \times 6.3 \]

\[ a = 0.345 \]

\[ \boxed{b = 3.59} \]

The Relation: \[ s = 0.34 + 3.59 t \]
(2) Fit the following data to \( P = \exp\left[ a + \frac{1}{bx} \right] \)

\[
\begin{align*}
X &: 22 \quad 23 \quad 24 \quad 25 \quad 26 \\
Y &: 0.368 \quad 0.223 \quad 0.134 \quad 0.082 \quad 0.05
\end{align*}
\]

**Solution**

Transform relation to 5th line form:

\[
\ln P = a + \frac{1}{bx} \quad \Rightarrow \quad Y = \ln P, \quad X = \frac{1}{x}
\]

\[
A_0 = a, \quad A_1 = \frac{1}{b}
\]

\[
N = 5, \quad \Sigma X_i = 0.2091, \quad \Sigma Y_i = -10.00
\]

\[
\begin{align*}
\Sigma X_i^2 &= 8.77 \times 10^{-3} \quad \Sigma XY = -0.4097 \\
5XY &= \Sigma XY - \frac{1}{n} (\Sigma X)(\Sigma Y) \\
&= -0.4097 - \frac{1}{5} (0.2091)(-10.007) \\
&= 8.79 \times 10^{-3}
\end{align*}
\]

\[
\begin{align*}
5XXY &= \Sigma X_i^2 - \frac{1}{n} (\Sigma X)^2 \\
&= 8.77 \times 10^{-3} - \frac{1}{5} (0.2091)^2 \\
&= 2.54 \times 10^{-5}
\end{align*}
\]

\[
A_0 = \frac{5XY}{5XX} = \frac{8.79 \times 10^{-3}}{2.54 \times 10^{-5}} = 346 \times 10^2 = 346.
\]

\[
\bar{X} = 0.0418, \quad \bar{Y} = -2.0014
\]

\[
A_1 = \frac{5XY}{5XXXY} = \frac{8.79 \times 10^{-3}}{2.54 \times 10^{-5}} = 346
\]
\[
A_1 \approx 3.46, \quad A_0 = -16.5
\]

\[
A_1 = \frac{1}{b} \implies b = 2.89 \times 10^{-3}
\]

\[
P = \exp \left[ -16.5 + \frac{1}{2.89 \times 10^{-3}} \right]
\]

(3) Fit the data in (1) above to a straight line that passes through the origin.

\[
y = mx
\]

\[
\frac{\partial \sum (y_i - \hat{y})^2}{\partial m} = 0 \implies \sum (y_i - mx_i)^2 = 0
\]

\[
2 \sum (y_i - mx_i)(x_i) = 0 \implies \sum y_i x_i = m \sum x_i^2
\]

\[
m = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{1240.7}{342} = 3.628
\]

\[
\text{Relation is } y = 3.628 x
\]
The least square Parabola:

The least square parabola approximating the set of pts. \((x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\) has the eqn.

\[ y = a_0 + a_1 x + a_2 x^2 \]

where the constants \(a_0, a_1, \text{ and } a_2\) are determined by solving simultaneously the eqns.

\[
\begin{align*}
\sum y &= a_0 \sum x + a_1 \sum x^2 + a_2 \sum x^2 \\
\sum xy &= a_0 \sum x + a_1 \sum x^2 + a_2 \sum x^3 \\
\sum x^2 y &= a_0 \sum x^2 + a_1 \sum x^3 + a_2 \sum x^4
\end{align*}
\]

Called the normal eqns. for the Least square Parabola.

This technique can be extended to obtain normal eqns. for Cubic and quartic curves.
The Least Square Parabola:

The least square parabola approximating the set of pts. \((x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\) has the eqn.

\[ y = a_0 + a_1 x + a_2 x^2 \]

where the constants \(a_0\), \(a_1\), and \(a_2\) are determined by solving simultaneously the eqns.

\[
\begin{align*}
\sum y &= a_0 \sum x + a_1 \sum x^2 + a_2 \sum x^3 \\
\sum xy &= a_0 \sum x + a_1 \sum x^2 + a_2 \sum x^3 \\
\sum x^2 y &= a_0 \sum x^2 + a_1 \sum x^3 + a_2 \sum x^4
\end{align*}
\]

Called the normal eqns. for the least square parabola.

\(\star\) This technique can be extended to obtain normal eqns. for cubic and quartic curves.
Example: Fit the following data to an equation of the form \( y = a_0 + a_1 x + a_2 x^2 \), by the method of least squares.

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>x^2</th>
<th>x^3</th>
<th>x^4</th>
<th>xy</th>
<th>x^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>157</td>
<td>100</td>
<td>1000</td>
<td>10000</td>
<td>1570</td>
<td>10000</td>
</tr>
<tr>
<td>20</td>
<td>179</td>
<td>400</td>
<td>8000</td>
<td>160000</td>
<td>3580</td>
<td>160000</td>
</tr>
<tr>
<td>30</td>
<td>210</td>
<td>900</td>
<td>27000</td>
<td>810000</td>
<td>6300</td>
<td>810000</td>
</tr>
<tr>
<td>40</td>
<td>252</td>
<td>1600</td>
<td>64000</td>
<td>4096000</td>
<td>10080</td>
<td>4096000</td>
</tr>
<tr>
<td>50</td>
<td>302</td>
<td>2500</td>
<td>125000</td>
<td>6250000</td>
<td>15100</td>
<td>6250000</td>
</tr>
<tr>
<td>60</td>
<td>361</td>
<td>3600</td>
<td>216000</td>
<td>12960000</td>
<td>21660</td>
<td>12960000</td>
</tr>
</tbody>
</table>

\[ \sum x = 0 \]  
\[ \sum x = 70 \]

Using least squares method to obtain the normal eqns. for 2nd order polynomial (Parabola).

\[ \sum y = a_0 N + a_1 \sum x + a_2 \sum x^2 \]
\[ \sum xy = a_0 \sum x + a_1 \sum x^2 + a_2 \sum x^3 \]
\[ \sum x^2 y = a_0 \sum x^2 + a_1 \sum x^3 + a_2 \sum x^4 \]

Subst. to obtain:

1. \[ 1461 = 6 a_0 + 70 a_2 \]
2. \[ 1431 = 70 a_0 + 252 a_2 \] \( \Rightarrow a_1 = 20.44 \)
3. \[ 17741 = 70 a_0 + 1414 a_2 \]

Eqn. 1 \( \times 70 \) - Eqn. 3 \( \times 6 \)
12270 = 420 \ a_0 + 4900 \ a_2 \\
106446 = 420 \ a_0 + 8484 \ a_2 \\
\[ 4176 = -3584 \ a_2 \quad \Rightarrow \quad a_2 = 1.165 \]
\[ \therefore a_0 = 229.9 \]

**Correlation** is a measure of the association between two random variables, both variables are assumed to be varying randomly. We do assume for this analysis that X and Y are related linearly, so the usual correlation coefficient gives a measure of the linear association between X and Y.

\[
R_{xy} = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}}
\]

for perfect correlation \( R = \pm 1 \) 

If there is no systematic relation between \( X \) and \( Y \) at all, \( R_{xy} = 0 \).
This fig. illustrate various correlation coefficients.

(a) $r_{xy} = 1$  
(b) $r_{xy} = -1$  
(c) $r_{xy} = 0$  
(d) $r_{xy} \approx 1$
It is required to fit the following equations to a straight line; so determine the constants, then calculate the correlation coefficient (Vey).

A. \[ y = ax^e + bx^2 - 2 \]

\[
\begin{align*}
x & : 1 & 2 & 3 & 4 & 5 \\
y & : 37.5 & 32.0 & 25.8 & 28.6 & 37.6
\end{align*}
\]

B. \[ y = ax^e + bx^e \]

\[
\begin{align*}
x & : 0 & 0.2 & 0.4 & 0.6 & 0.8 & 1.0 \\
y & : -1.12 & 0.026 & 1.15 & 2.32 & 3.59 & 5.0
\end{align*}
\]

C. \[ \ln y = ax^e + bx \]

\[
\begin{align*}
x & : 0.21 & 0.27 & 0.35 & 0.38 & 0.43 \\
y & : 10 & 22 & 70 & 100 & 240
\end{align*}
\]
\[ y = \frac{x}{a+bx} \]

\[ y: 3.5 \quad 7.2 \quad 12.6 \quad 16.4 \quad 20.2 \]

\[ x: 100 \quad 200 \quad 300 \quad 400 \quad 500 \]

\[ C^2 = \frac{C_i^2}{2C_iKT+1} \]

\[ C: 2.5 \quad 1.65 \quad 1.18 \quad 0.95 \quad 0.88 \]

\[ t: 10 \quad 15 \quad 20 \quad 25 \quad 30 \]

\[ K = \frac{-E}{RT} \]

\[ K: 1.22 \quad 2.72 \quad 4.95 \quad 7.39 \quad 11.0 \]

\[ T: 316.46 \quad 322.58 \quad 331.16 \quad 336.7 \quad 340.1 \]