9.12 Model II: principal axis regression

Example 9.7 The lower arm and lower leg length (cm) of a small cohort of female undergraduates

In an investigation into human morphology, the lower arm and lower leg lengths of 12 female undergraduates were recorded (Table 9.15) on one particular day.

Table 9.15 Lower arm and lower leg length (cm) in a small cohort of female undergraduates

<table>
<thead>
<tr>
<th>Student</th>
<th>Lower arm length (cm)</th>
<th>Lower leg length (cm)</th>
<th>Student</th>
<th>Lower arm length (cm)</th>
<th>Lower leg length (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>24.0</td>
<td>39.0</td>
<td>7</td>
<td>26.0</td>
<td>41.0</td>
</tr>
<tr>
<td>2</td>
<td>25.0</td>
<td>40.0</td>
<td>8</td>
<td>25.5</td>
<td>40.0</td>
</tr>
<tr>
<td>3</td>
<td>23.0</td>
<td>36.0</td>
<td>9</td>
<td>24.2</td>
<td>36.7</td>
</tr>
<tr>
<td>4</td>
<td>26.0</td>
<td>43.0</td>
<td>10</td>
<td>25.0</td>
<td>38.0</td>
</tr>
<tr>
<td>5</td>
<td>24.0</td>
<td>38.0</td>
<td>11</td>
<td>24.0</td>
<td>41.0</td>
</tr>
<tr>
<td>6</td>
<td>23.0</td>
<td>35.0</td>
<td>12</td>
<td>24.5</td>
<td>40.0</td>
</tr>
</tbody>
</table>

BOX 9.8 How to carry out a Model II principal axis regression: drawing a regression line

1. Have the criteria for using this test been met?
   Yes (9.12.2.ii.).

2. How to work out $b_1$
   i. First decide which of the variables to call $x$ and which to call $y$. Let the measurements of arm length (cm) be $x$ and the measurements of leg length (cm) be $y$ (Fig. 9.16).
   ii. Calculate the basic terms.

To carry out this regression analysis, you first need to calculate the following for the two samples: Means ($\bar{x}$ and $\bar{y}$).

\[
\sum x = 24.0 + 25.0 + 23.0 + \ldots + 25.0 + 24.0 + 24.5 = 294.2
\]
\[
\bar{x} = \frac{294.2}{12} = 24.51666
\]

\[
\sum y = 39.0 + 40.0 + 36.0 + \ldots + 38.0 + 41.0 + 40.0 = 467.7
\]
\[
\bar{y} = \frac{467.7}{12} = 38.975
\]

Sum of $x$ values ($\Sigma x$), sum of $y$ values ($\Sigma y$).

\[
\sum x = 24.0 + 25.0 + 23.0 + \ldots + 25.0 + 24.0 + 24.5 = 294.2
\]

(continued)
The variances \((s_x^2\) and\(s_y^2\)):

\[
\sum (x)^2 = (294.2)^2 = 86553.64
\]
\[
\frac{\sum x^2}{n} = \frac{86553.64}{12} = 7212.80333
\]
\[
\sum (x^2) = 24.0^2 + 25.0^2 + 23.0^2 + \ldots + 25.0^2 + 24.0^2 + 24.5^2 = 576 + 625 + 529 + \ldots + 625 + 576 + 600.25 = 7224.14
\]
\[
s_x^2 = \frac{7224.14 - 7212.80333}{12.0 - 1.0} = \frac{11.33667}{11.0} = 1.03061
\]
\[
\sum y^2 = (467.7)^2 = 218743.29
\]
\[
\frac{\sum y^2}{n} = \frac{218743.29}{12} = 18228.6075
\]
\[
\sum (y^2) = 39.0^2 + 40.0^2 + 36.0^2 + \ldots + 38.0^2 + 41.0^2 + 40.0^2 = 1521 + 1600 + 1296 + \ldots + 1444 + 1681 + 1600 = 18287.89
\]
\[
s_y^2 = \frac{18287.89 - 18228.6075}{12 - 1} = \frac{59.2925}{11} = 5.38932
\]

The number of pairs of observations \(n = 12\).
These calculations are summarized in Table 9.16.

**Table 9.16** Calculating a Model II principal axis regression using the data from Example 9.7. The lower arm and leg length (cm) in a small cohort of female undergraduates

<table>
<thead>
<tr>
<th>Length of arm (cm) (x)</th>
<th>Length of leg (cm) (y)</th>
<th>(x \times y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>24.00</td>
<td>39.00</td>
<td>936.00</td>
</tr>
<tr>
<td>25.00</td>
<td>40.00</td>
<td>1000.00</td>
</tr>
<tr>
<td>23.00</td>
<td>36.00</td>
<td>828.00</td>
</tr>
<tr>
<td>26.00</td>
<td>43.00</td>
<td>1118.00</td>
</tr>
<tr>
<td>24.00</td>
<td>38.00</td>
<td>912.00</td>
</tr>
<tr>
<td>23.00</td>
<td>35.00</td>
<td>805.00</td>
</tr>
<tr>
<td>26.00</td>
<td>41.00</td>
<td>1066.00</td>
</tr>
<tr>
<td>25.50</td>
<td>40.00</td>
<td>1020.00</td>
</tr>
<tr>
<td>24.20</td>
<td>36.70</td>
<td>888.14</td>
</tr>
<tr>
<td>25.00</td>
<td>38.00</td>
<td>950.00</td>
</tr>
<tr>
<td>24.00</td>
<td>41.00</td>
<td>984.00</td>
</tr>
<tr>
<td>24.50</td>
<td>40.00</td>
<td>980.00</td>
</tr>
<tr>
<td>(n = 12)</td>
<td>(n = 12)</td>
<td>(n = 12)</td>
</tr>
<tr>
<td>(\Sigma x = 294.2)</td>
<td>(\Sigma y = 467.70)</td>
<td>(\Sigma(xy) = 11487.14)</td>
</tr>
<tr>
<td>(\bar{x} = 24.51667)</td>
<td>(\bar{y} = 38.97500)</td>
<td></td>
</tr>
</tbody>
</table>

(continued)
iii. Calculate the variance for products \(xy\), \(s_{xy}^2\), where

\[
s_{xy}^2 = \frac{SP(xy)}{n-1}
\]

\[
SP(xy) = \sum xy - \frac{\sum x \sum y}{n}
\]

\[
\sum (xy) = (24.0 \times 39.0) + (25.0 \times 40.0) + \ldots + (24.0 \times 41.0) + (24.5 \times 40.0)
\]

\[
= 936.00 + 1000.00 + \ldots + 984.00 + 980.00 = 11487.14
\]

\[
SP(xy) = 11487.14 - \frac{(294.2 \times 467.7)}{12} = 11487.14 - \frac{137597.34}{12} = 11487.14 - 11466.445 = 20.695
\]

\[
s_{xy}^2 = \frac{SP(xy)}{n-1} = \frac{20.695}{11} = 1.88136
\]

iv. The slope of the principal axis \(b_1\) is calculated as

\[
b_1 = \frac{s_y^2 - s_x^2 + \sqrt{\left(s_y^2 - s_x^2\right)^2 + 4\left(s_{xy}^2\right)^2}}{2s_{xy}^2}
\]

\[
First calculate the components of this equation.
\]

\[
s_y^2 - s_x^2 = 5.38932 - 1.03061
\]

\[
= 4.35871
\]

\[
\left(s_y^2 - s_x^2\right)^2 = (4.35871)^2
\]

\[
= 18.99837
\]

\[
4\left(s_{xy}^2\right)^2 = 4 \times (1.88136)^2
\]

\[
= 14.15812
\]

\[
2s_{xy}^2 = 2 \times 1.88136
\]

\[
= 3.76273
\]

\[
\text{Therefore } b_1 = \frac{4.35871 + \sqrt{18.99837 + 14.15812}}{3.76227}
\]

\[
b_1 = \frac{4.35871 + \sqrt{33.15649}}{3.76227} = \frac{4.35871 + 5.75817}{3.76227} = \frac{10.11688}{3.76227} = 2.68871
\]

3. What is the regression equation?

The equation for the principal axis is \(y = \bar{y} + b_1(x - \bar{x})\). Using the values you have already calculated:

\[
y = \bar{y} + b_1(x - \bar{x}) = 38.975 + 2.68871(x - 24.51667)
\]

\[
= 38.975 + 2.68871x - 65.91822 = -26.94319 + 2.68871x
\]
**4. How to draw the line**

In the same way as other regressions you need to calculate the $x$ and $y$ values for three points using $x$ values that lie within your data set, and your regression equation. Your line then passes through these points within the range of your data set. This is the regression line and if you add it to your figure it should be labelled with the regression equation.

The full regression equation for this example is

$$y = -26.94319 + 2.68871x$$

Therefore if

- $x = 23$ cm,
  $$y = -26.94319 + (2.68871 \times 23) = 34.9$$ cm
- and if $x = 24.5$ cm then $y = 38.9$ cm, and if $x = 26$ cm then $y = 43$ cm.

The data from Table 9.15 and the regression line are shown on Fig. 9.15.

---

**5. Testing the significance of the line**

The principal axis regression allows you to identify a regression line that reflects the linear nature of the association between your two variables. However, unlike the simple linear regression there is no simple associated process for testing a hypothesis relating to this association. Therefore you should also use a correlation analysis to confirm the strength and significance of the association between the two variables.

An examination of the data (see Interactive Exercises 1) indicates that it is non-parametric, therefore a two-tailed Spearman rank correlation was carried out. From this it was concluded that there is a moderate ($r_s = 0.725$) statistically significant ($p = 0.05$) positive correlation between the lower arm and lower leg length in a cohort of female students which can be described by

$$y = -26.94319 + 2.68871x$$ (Fig. 9.15).