**LEAST SQUARE ESTIMATION**

Lets us look at an example using sample data to estimate least squares (and in particular 01)

Approach

Assume a sample of n subjects, observing y value of the response variable and the predictor variable ‘x’.

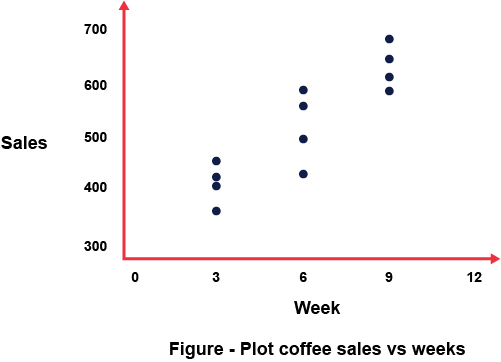
Consider the following dataset, where we have coffee sales for 12 weeks [Table 1].

The idea is to choose the estimates for01 that best fits the data. Remember the regression model “fitted equation”:

|  |  |
| --- | --- |
| Shelf space (x) | Sales (y) [Weekly] |
| 6 | 526 |
| 6 | 434 |
| 6 | 581 |
| 3 | 421 |
| 9 | 630 |
| 3 | 412 |
| 9 | 560 |
| 3 | 443 |
| 9 | 590 |
| 6 | 570 |
| 3 | 346 |
| 9 | 672 |

Table 1 – Coffee Sales Data

Experiment was conducted over a period of 12 weeks [4 weeks with 3 foot shelf space, 4 weeks with 6 foot and 4 weeks with 9 foot shelf space]



By choosing estimates for 01, we aim to minimize the distance of the data points from the fitted line.

For each response value yi, with a predictor variable x; we obtain a fitted line, we choose 01, (as estimates for 01)

x

**REMEMBER the aim is to minimize the sum of squared error as we saw in our lecture notes.**

Note: it is probably really easy to do this in Python, Excel or R, but we want to understand mathematically what is going on so let us continue.

From 2, we need to only know what is: so then we can rewritten as

1is estimated as:

= =

o is estimated as:

=

= -

Now that might seem like a lot to calculate, but a few shortcut equations can help us calculate this easily:

=

=. -

All that is left to do is calculate the values and plug it back into the line equation –

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Week | Shelf Space (x) | Weekly Sales (y) |  | xy |  |
| 1 | 6 | 526 | 36 | 3156 | 276676 |
| 2 | 3 | 421 | 9 | 1263 | 177241 |
| 3 | 6 | 581 | 36 | 3486 | 337561 |
| 4 | 9 | 630 | 81 | 5670 | 396900 |
| 5 | 3 | 412 | 9 | 1236 | 169744 |
| 6 | 9 | 560 | 81 | 5040 | 313600 |
| 7 | 6 | 434 | 36 | 2604 | 188356 |
| 8 | 3 | 443 | 9 | 1329 | 195249 |
| 9 | 9 | 590 | 81 | 5310 | 348100 |
| 10 | 6 | 570 | 36 | 3420 | 324900 |
| 11 | 3 | 346 | 9 | 1038 | 119716 |
| 12 | 9 | 672 | 81 | 6048 | 451584 |

Also:

SSxx =

= 504 –

**SSxx = 72**

= –

**=** 3300627 **-**

To obtain **LEAST SQUARES ESTIMATE OF OUR LINEAR REGRESSION** relation (01,x)

**Finally**

The interpretation for this “fitted line” is that the “estimate for the increase in mean weekly sales, in account of increasing the shelf space by 1 foot is 34.583 bags of coffee”

COFFEE SALES 34.583 bags

For every 1 foot

In shelf space