Case study on Simple Linear Regression:

Coffee Sales and Shelf Space

Context:

**A marketer is interested in the effect of changing shelf height (x1) and shelf width (x2) on the weekly sales (y) of her brand of laundry detergent in a grocery store.**

Objective:

We have at least one variable that is known (in some cases it is controllable), and a response variable that is a random variable. We would like to fit a model that relates the response to the known or controllable variable(s).

A marketer is interested in the relation between the width of the shelf space for her brand of coffee (x) and weekly sales (y) of the product in a sub urban supermarket (assume the height is always at eye level). Marketers are well aware of the concept of ‘compulsive purchases’, and know that the more shelf space their product takes up, the higher the frequency of such purchases. She believes that in the range of 3 to 9 feet, the mean weekly sales will be linearly related to the width of the shelf space. Further, among weeks with the same shelf space, she believes that sales will be normally distributed with unknown standard deviation σ (that is, σ measures how variable weekly sales are at a given amount of shelf space). Thus, she would like to fit a model relating weekly sales y to the amount of shelf space x her product receives that week.

That is, she is fitting the model:

**Equation:**

y = β0 + β1x + ε, so that y|x ∼ N(β0 + β1x, σ).

**Steps**

We now have the problem of using sample data to compute estimates of the parameters β0 and β1. First, we take a sample of n subjects, observing values y of the response variable and x of the predictor variable. We would like to choose as estimates for β0 and β1, the values b0 and b1 that ‘best fit’ the sample data. Consider the coffee example mentioned earlier. Suppose the marketer conducted the experiment over a twelve week period (4 weeks with 3’ of shelf space, 4 weeks with 6’, and 4 weeks with 9’), and observed the sample data in Table 1.

Table 1: Coffee sales data for n = 12 weeks

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

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Week | Space(x) | Sales(y) | x^2 | xy | y^2 |
| 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 | 196249 |
| 9 | 9 | 590 | 81 | 5310 | 348100 |
| 10 | 6 | 570 | 36 | 3420 | 324900 |
| 11 | 3 | 346 | 9 | 1038 | 119716 |
| 12 | 9 | 672 | 81 | 6048 | 451584 |
|  | 72 | 6185 | 504 | 39600 | 3300627 |

For the coffee data, we observe the following summary statistics in Table 2.

Table 2: Summary Calculations — Coffee sales data

From this, we obtain the following sums of squares and crossproducts.

SSxx = Σ (x – x̅) 2 = Σ x2 − ( Σ x)2 /n = 504 − (72)2/12 = 72

SSxy = Σ (x − x̅)(y − ȳ) = Σ xy − ( Σ x)( Σ y) /n = 39600 − (72)(6185) /12 = 2490

SSyy = Σ (y − ȳ) 2 = Σ y2 − ( Σ y)2 /n = 3300627 − (6185)2 /12 = 112772.9

From these, we obtain the least squares estimate of the true linear regression relation (β0+β1x)

b1 = SSxy /SSxx = 2490 72 = 34.5833

b0 = Σ y /n − b1 Σ x /n = 6185 12 − 34.5833(72 /12 ) = 515.4167 − 207.5000 = 307.967.

yˆ = b0 + b1x = 307.967 + 34.583x

So the fitted equation, estimating the mean weekly sales when the product has x feet of shelf space is ˆy = βˆ 0 + βˆ 1x = 307.967 + 34.5833x. Our interpretation for b1 is “the estimate for the increase in mean weekly sales due to increasing shelf space by 1 foot is 34.5833 bags of coffee”. Note that this should only be interpreted within the range of x values that we have observed in the “experiment”, namely x = 3 to 9 feet.

Source: <http://users.stat.ufl.edu/~winner/qmb3250/notespart2.pdf>