An Example: Delivery Time Data [Montgomery et al., 2012]

Suppose that an industrial engineer was employed to analyze the product delivery and service operations for vending machines.

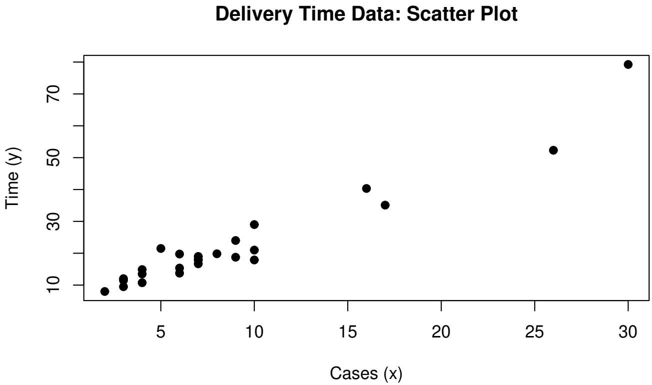
- He/She suspects that the time required by a delivery person to load and service a machine is related to the number of cases of product delivered.
- The engineer visits 25 randomly chosen vending machines, and noted the in-outlet delivery time (in minutes) and the volume of product delivered (in cases).
- The question he/she is trying to answer: how much time is needed for the given number of cases?

An Example: Delivery Time Data

Table: Time Delivery Data

Obs.	Del. Time	Cases	Obs.	Del. Time	Cases
1.	16.68	7	14.	19.75	6
2.	11.5	3	15.	24	9
3.	12.03	3	16.	29	10
4.	14.88	4	17.	15.35	6
5.	13.75	6	18.	19	7
6.	18.11	7	19.	9.5	3
7.	8	2	20.	35.1	17
8.	17.83	7	21.	17.9	10
9.	79.24	30	22.	52.32	26
10.	21.5	5	23.	18.75	9
11.	40.33	16	24.	19.83	8
12.	21	10	25.	10.75	4
13.	13.5	4			

Scatter Plot of Delivery Time Data



Least Squares Estimation

- Consider a data set consists of n points (x_i, y_i) $i = 1, \ldots, n$. Here, x_i is an independent variable and y_i is a dependent variable.
- The model function $f(\boldsymbol{x}, \boldsymbol{\beta})$ depends on $\boldsymbol{x} = (\mathrm{x}_1, \ldots, \mathrm{x}_p)$ (p–regressors) and m unknown parameters $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)$.
- The goal is to find the parameter values for the model that “best"fits the data.

Least Squares Estimation

- The goodness of model fit is measured by its residual, defined as the difference between the actual value of the dependent variable and the value predicted by the model

$$e_i = y_i - f(x_i, \boldsymbol{\beta}).$$

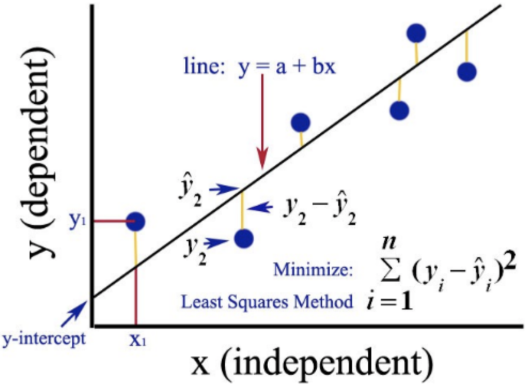
- The least-squares method finds the optimal parameter values by minimizing the sum of squared residuals

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{n} e_i^2.$$

Linear Least Squares

- Suppose we wish to fit a straight line. Let β_0 and β_1 denote the intercept and the slope, respectively. The model function takes the form
$$f(x, \boldsymbol{\beta}) = \beta_0 + \beta_1 x$$
- Optimal values of β_0 and β_1 are found by minimizing
$$S(\boldsymbol{\beta}) = \sum_{i=1}^{n} \{y_i - (\beta_0 + \beta_1 x_i)\}^2.$$

Ordinary Least Squares



Note: The figure is taken from https://medium.com/analytics-vidhya

Definitions

Let A be a $n \times n$ symmetric matrix, then A is:

- **positive definite** if $\mathbf{x}'\mathrm{A}\mathbf{x} > 0$ for all $\mathbf{x} \neq 0$ in $\mathbb{R}^n$.
- **negative definite** if $\mathbf{x}'\mathrm{A}\mathbf{x} < 0$ for all $\mathbf{x} \neq 0$ in $\mathbb{R}^n$.
- **positive semidefinite** if $\mathbf{x}'\mathrm{A}\mathbf{x} \geq 0$ for all $\mathbf{x} \neq 0$ in $\mathbb{R}^n$.
- **negative semidefinite** if $\mathbf{x}'\mathrm{A}\mathbf{x} \leq 0$ for all $\mathbf{x} \neq 0$ in $\mathbb{R}^n$.
- **indefinite** if $\mathbf{x}'\mathrm{A}\mathbf{x} > 0$ for some $\mathbf{x}$ in $\mathbb{R}^n$ and $\mathbf{x}'\mathrm{A}\mathbf{x} < 0$ for some other $\mathbf{x}$ in $\mathbb{R}^n$.

Maxima/Minima

Given a multivariable function $f(\mathbf{x})$ (where $\mathbf{x} = (x_1, \ldots, x_n)'$) defined as $f : D(\subset \mathbb{R}^n)$ onto $\mathbb{R}$, we aim to find

$$\min_{\mathbf{x}\in D} f(\mathbf{x})$$

or

$$\max_{\mathbf{x}\in D} f(\mathbf{x}).$$

Definitions

A point $\mathbf{x}^*$ is:

- a **max** of f in D if $f(\mathbf{x}^*) \geq f(\mathbf{x})$ for all $\mathbf{x} \in D$;
- a **strict max** of f in D if it is a max and $f(\mathbf{x}^*) > f(\mathbf{x})$ for all $\mathbf{x} \neq \mathbf{x}^*$;
- a **local max** of f if there is a neighborhood of $\mathbf{x}^*$, $B_r(\mathbf{x}^*)$ such that $f(\mathbf{x}^*) \geq f(\mathbf{x})$ for all $\mathbf{x} \in B_r(\mathbf{x}^* \cap D)$;
- a **strict local max** of f if there is a neighborhood of $\mathbf{x}^*$, $B_r(\mathbf{x}^*)$ such that $f(\mathbf{x}^*) > f(\mathbf{x})$ for all $\mathbf{x} \in B_r(\mathbf{x}^* \cap D)$.

Definitions

A point $\mathbf{x}^*$ is:

- a **min** of f in D if $f(\mathbf{x}^*) \leq f(\mathbf{x})$ for all $\mathbf{x} \in D$;
- a **strict min** of f in D if it is a max and $f(\mathbf{x}^*) < f(\mathbf{x})$ for all $\mathbf{x} \neq \mathbf{x}^*$;
- a **local min** of f if there is a neighborhood of $\mathbf{x}^*$, $B_r(\mathbf{x}^*)$ such that $f(\mathbf{x}^*) \leq f(\mathbf{x})$ for all $\mathbf{x} \in B_r(\mathbf{x}^* \cap D)$;
- a **strict local min** of f if there is a neighborhood of $\mathbf{x}^*$, $B_r(\mathbf{x}^*)$ such that $f(\mathbf{x}^*) < f(\mathbf{x})$ for all $\mathbf{x} \in B_r(\mathbf{x}^* \cap D)$.

First order condition

Let $f : D \subset \mathbb{R}^n \to \mathbb{R}$ be a real–valued function of n variables. If a point $\mathbf{x}^*$ is a local max or min of f in D and if $\mathbf{x}^*$ is an interior point of D, then

$$\frac{\partial f}{\partial x_i} = 0 \;\; \text{for}, i = 1, \ldots, n$$

in other words the gradient vanishes at $\mathbf{x}^*$.

First order condition

Definition 1

A point $\mathbf{x}^*$ is a **critical point** of a function $f(x_1, \ldots, x_n)$ if it satisfies

$$\frac{\partial f}{\partial x_i} = 0 \text{ for, } i = 1, \ldots, n$$

Remark 0.1 (Hessian Matrix)

In case of more than two variables, local minima/maxima is decided using Hessian matrix. The (i, j) entry of a typical Hessian matrix is

$$(H_f)_{i,j} = \frac{\partial^2 f(x_1, \dots, x_n)}{\partial x_i \partial x_j}.$$

Let x_0 be a critical point of f.

1. If Hessian matrix is positive definite at $(x_1, \dots, x_n) = \boldsymbol{x} = \boldsymbol{x}_0$ then f attains a strict local minima at $\boldsymbol{x}_0$.
2. If Hessian matrix is negative definite at $\boldsymbol{x} = \boldsymbol{x}_0$ then f attains a strict local maxima at $\boldsymbol{x}_0$.
3. Otherwise $\boldsymbol{x}_0$ is a **saddle point**.

Remark 0.2 (Sylvester's criterion)

The real-symmetric matrix H is positive definite if and only if all the leading principal minors of H are positive.

In other words all the following matrices have a positive determinant:

1. The upper left 1×1 corner of H,
2. The upper left 2×2 corner of H,

$\vdots$

n. H itself.

Linear Least Squares Estimators (LSEs)

Let us go back to our least squares problem.

$$\frac{\partial S(\boldsymbol{\beta})}{\partial \beta_0} = -2\sum_{i=1}^{n}\{y_i - (\beta_0 + \beta_1 x_i)\} = 0 \tag{1}$$

$$\frac{\partial S(\boldsymbol{\beta})}{\partial \beta_1} = -2\sum_{i=1}^{n}\{y_i - (\beta_0 + \beta_1 x_i)\}x_i = 0 \tag{2}$$

Solving (1) and (2) for β_0 and β_1, we get the LSEs of (β_0, β_1) as

$$\widehat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \text{ and } \widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1\overline{x}, \tag{3}$$

where $\overline{y} = \sum_{i=1}^{n} y_i/n$, $\overline{x} = \sum_{i=1}^{n} x_i/n$, $S_{xy} = \sum_{i=1}^{n} y_i(x_i - \overline{x})$ and $S_{xx} = \sum_{i=1}^{n}(x_i - \overline{x})^2$.

Linear Least Squares Estimators (LSEs)

Second derivative test:

$$\frac{\partial^2 S(\boldsymbol{\beta})}{\partial \beta_0^2} = 2n, \ \frac{\partial^2 S(\boldsymbol{\beta})}{\partial \beta_1^2} = 2\sum_{i=1} x_i^2 \text{ and } \frac{\partial^2 S(\boldsymbol{\beta})}{\partial \beta_0 \partial \beta_1} = 2\sum_{i=1} x_i.$$

$$\frac{\partial^2 S(\boldsymbol{\beta})}{\partial \beta_0^2}\frac{\partial^2 S(\boldsymbol{\beta})}{\partial \beta_1^2} - \left[\frac{\partial^2 S(\boldsymbol{\beta})}{\partial \beta_0 \beta_1}\right]^2 = 4n\sum_{i=1}^{n}(x_i - \overline{x})^2 > 0 \quad (4)$$

and

$$\frac{\partial^2 S(\boldsymbol{\beta})}{\partial \beta_0^2} = 2n > 0. \quad (5)$$

From (4) and (5) it is concluded that $(\widehat{\beta}_0, \widehat{\beta}_1)$ given in (3) minimized the objective function.

References

Montgomery, D. C., Peck, E. A., and Vining, G. G. (2012).
Introduction to Linear Regression Analysis.
John Wiley Sons, Hoboken, New Jersey, 5th edition.