Basic Math Review

Some Greek letters

Greek letters are frequently used to denote either specific population properties, to name specific statistical test or as mathematical operator.

Greek Letter	Phonetic	Usage
$\boldsymbol{\alpha}$	*alpha*	error of first type
$\boldsymbol{\beta}$	*beta*	error of second type / regression parameters
$\boldsymbol{\varepsilon}$	*epsilon*	regression population error term
$\boldsymbol{\gamma}$	*gamma*	parameter in some distributions and functions
$\boldsymbol{\lambda}$	*lambda*	parameter in some distributions and functions
$\boldsymbol{\mu}$	*mu*	expected population mean
$\boldsymbol{\pi}$	*pi*	population probability in binomial distribution
$\boldsymbol{\rho}$	*rho*	population correlation coefficient
$\boldsymbol{\sigma}$	*sigma*	population standard deviation
$\boldsymbol{\chi}$	*chi*	χ^2-test
$\boldsymbol{\Pi}$	*capital pi*	multiplication symbol
$\boldsymbol{\Sigma}$	*capital sigma*	summation symbol

Standard Symbols and Definition

Operation	Meaning
$\frac{\boldsymbol{Numerator}}{\boldsymbol{Denominator}}$	ratio between the numerator and the denominator
$\times$ ***or*** $\cdot$, ***and*** $\div$ ***or*** $/, +, -$	multiplication and division take precedence over addition and subtraction
$\boldsymbol{X < Y}$	X is less than Y
$\boldsymbol{X \leq Y}$	X is less or equal than Y
$\boldsymbol{X \pm Y}$	X plus minus Y, i.e., the two values $X + Y$ and $X - Y$
$\lvert\boldsymbol{X}\rvert$	The absolute value $\lvert X\rvert = \begin{cases} X & \text{for } X \geq 0 \\ -X & \text{for } X < 0 \end{cases}$
$\frac{\boldsymbol{1}}{\boldsymbol{X}} = \boldsymbol{X^{-1}}$	Reciprocal of X
$\boldsymbol{X^n}$	The n^{th} power of X; for n being integer we get $x^n = \underbrace{x \cdot x \cdots x}_{n-\text{times}}$
$\sqrt{\boldsymbol{X}} = \boldsymbol{X^{\frac{1}{2}}}$	square root of X
$\boldsymbol{i} \in \{\boldsymbol{1, 2, \dots n}\}$	i is an element in the set $\{\boldsymbol{1, 2, \dots n}\}$. It takes the values $\boldsymbol{1, 2, \dots n.}$

Notation of Variables

A variable is denoted by a capital letter X while a lower case letter with a subscript index x_i relates to a specific observation. The index i ranges from $1, 2, \dots, n$. The number of observations in a variable is n. Therefore, $X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$. For example, if X has $n = 4$ observation then $X = \begin{pmatrix} x_1 = 3 \\ x_2 = 2 \\ x_3 = 5 \\ x_4 = 4 \end{pmatrix}$.

Ranked Data

- Statisticians frequently work with an ascending sorted sequence of observations which is denoted by square brackets $X_{[ranked]} = \begin{pmatrix} x_{[1]} \\ x_{[2]} \\ \vdots \\ x_{[n]} \end{pmatrix}$. For example, $X_{[ranked]} = \begin{pmatrix} x_{[1]} = 2 \\ x_{[2]} = 3 \\ x_{[3]} = 4 \\ x_{[4]} = 5 \end{pmatrix}$ with $x_{[1]} \leq x_{[2]} \leq x_{[3]} \leq x_{[4]}$.
- Should two observations have the same rank, such as $x_i = 5$ and $x_j = 5$, then the ranks $[r]$ and $[r+1]$ will be assigned arbitrarily, that is, $x_{[r]} = 5$ and $x_{[r+1]} = 5$, respectively.

Basic Summation $\sum$ −Rules:

- $\sum_{i=1}^{n} x_i \equiv x_1 + x_2 + \cdots + x_n$. The lower index $i = 1$ express the starting value of the summation sequence and the upper index n the value where the summation index i stops.
- more specifically $\sum_{i=2}^{5} x_i = x_2 + x_3 + x_4 + x_5$
- for a sum over a constant c we get $\sum_{i=1}^{n} c = n \cdot c$
- for a mixture of a constant and a variable $\sum_{i=1}^{n} c \cdot x_i = c \cdot \sum_{i=1}^{n} x_i$
- for an additive mixture of variables $\sum_{i=1}^{n} (x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$
- Inequalities (so do not confuse either side of the expression; they lead to different results):

  $\sum_{i=1}^{n} x_i \cdot y_i \neq \sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i$

  $\sum_{i=1}^{n} x_i^2 \neq \left(\sum_{i=1}^{n} x_i\right)^2$

- Special rule used when we are dealing with ranks: $\sum_{i=1}^{n} i = \frac{n}{2}\cdot(n+1)$
- Doubly index variables x_{ij} in a data table with *I* rows for the observations and *J* columns for the variables:

  Let: $\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1J} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2J} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{ij} & \cdots & x_{iJ} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{I1} & x_{I2} & \cdots & x_{Ij} & \cdots & x_{IJ} \end{bmatrix}$

  Then the i^{th} row sum is $x_{i+} = \sum_{j=1}^{J} x_{ij}$ and the j^{th} column sum is $x_{+j} = \sum_{i=1}^{I} x_{ij}$ and the total sum becomes $x_{++} = \sum_{i=1}^{I}\sum_{j=1}^{J} x_{ij} = \sum_{i=1}^{I} x_{i+} \text{ or } \sum_{j=1}^{J} x_{+j}$
- Applications (not test relevant):
  - Rewrite: $\sum_{i=1}^{n}(x_i + y_i)^2 = \sum_{i=1}^{n}\left(x_i^2 + 2\cdot x_i\cdot y_i + y_i^2\right) = \sum_{i=1}^{n} x_i^2 + 2\cdot\sum_{i=1}^{n} x_i\cdot y_i + \sum_{i=1}^{n} y_i^2$
- The R function for summations:
  - **`sum()`** calculated the sum over the elements of a vector
  - **`rowSums()`** calculates along the rows of a matrix a vector of row sums.
  - **`colSums()`** calculates along the columns of a matrix a vector of column sums.

Finding the minimum of a quadratic function (not test relevant)

- In statistic, we encounter frequently the need to find an optimal value of an estimation function, such as for the central tendency.
- The minimum is found at that point, where the slope of the function is zero. The slope of a function is measured by the first derivative.
- Basic rules of derivatives:

  $$\frac{\partial}{\partial x} a \cdot x^n = a \cdot n \cdot x^{n-1} \quad \text{Example: } \frac{\partial}{\partial x} 3 \cdot x^2 = 6 \cdot x$$

  $$\frac{\partial}{\partial x}\left(f(x) + g(x)\right) = \frac{\partial}{\partial x} f(x) + \frac{\partial}{\partial x} g(x) \quad \text{Example: } \frac{\partial}{\partial x}\left(3 \cdot x^2 + 5 \cdot x^{-1}\right) = \underbrace{3 \cdot 2}_{=6} \cdot x \underbrace{+5 \cdot -1}_{=-5} \cdot x^{-2}$$

- <u>Example:</u> The Arithmetic Mean

  Which value of θ (theta) minimizes the quadratic expression $\sum_{i=1}^{n}(x_i - \theta)^2$?

  Analysis Steps:

  [a] Multiply the term under the square out:

  $$f(\theta) = \sum_{i=1}^{n}(x_i - \theta)^2 = \sum_{i=1}^{n} x_i^2 - 2 \cdot \theta \cdot \sum_{i=1}^{n} x_i + n \cdot \theta^2$$

  [b] Take the first derivative with regard to θ, which is the slope of $f(\theta)$ at θ:

  $$\frac{\partial}{\partial \theta}\left(\sum_{i=1}^{n} x_i^2 - 2 \cdot \theta \cdot \sum_{i=1}^{n} x_i + n \cdot \theta^2\right) = -2 \cdot \sum_{i=1}^{n} x_i + 2 \cdot n \cdot \theta$$

[c] At its maximum or minimum the first derivative (that is, the slope) is zero.

Therefore set the expression to zero:

$$-2\cdot\sum_{i=1}^{n} x_i + 2\cdot n\cdot\theta = 0$$

[d] Solve the expression for the unknown parameter θ:

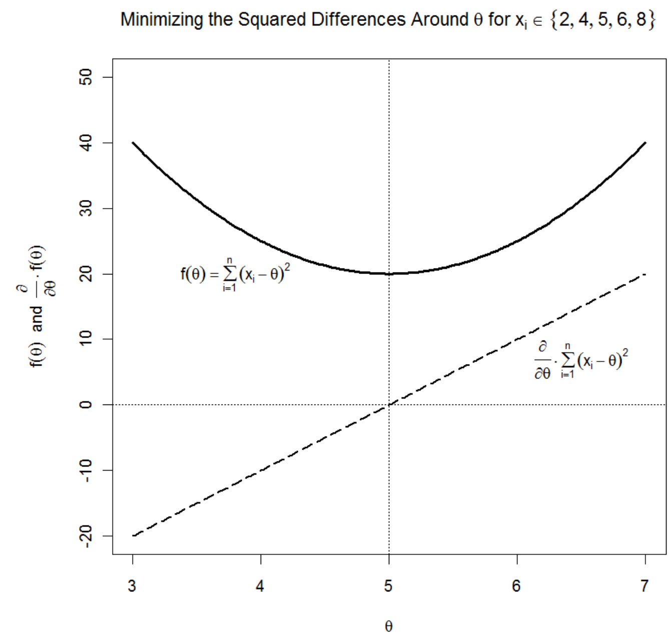
$$\theta = \frac{\sum_{i=1}^{n} x_i}{n}$$

$\Rightarrow$ This is the well-know arithmetic mean!!!

<u>Example:</u> The data values are $x_i \in \{2,5,4,6,8\}$.

Thus the function to be minimized with respect to θ is

$$f(\theta)=(2-\theta)^2+(5-\theta)^2+(4-\theta)^2+(6-\theta)^2+(8-\theta)^2$$

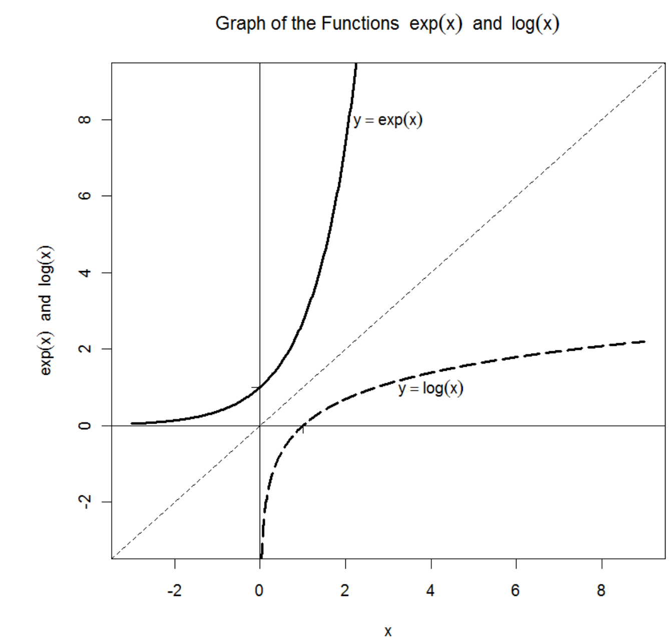


The solution is found at $\theta \Rightarrow \bar{x} = 5$.

The exponential and logarithmic functions (not test relevant)

- Both functions are inversely related: $x = \exp(log(x))$ and $x = \log(\exp(x))$.

  Note: the support (valid arguments for the variable x) of the logarithmic function is limited from below by zero, that is, $x \in]0, \infty]$.



- The both functions distort the constant distance units in the variable x.
  E.g., $\Delta x_1 = 10 - 1 = 9$ and
  $\Delta x_2 = 20 - 11 = 9$ but
  $\log(10) - \log(1) = 2.302585$
  but $\log(20) - \log(11) = 0.597837$, respectively.
- Basic rules:
  - Logarithmic function:
    $\log(x \cdot y) = \log(x) + \log(y)$,
    $\log(x/y) = \log(x) - \log(y)$ and
    $\log(x^y) = y \cdot \log(x)$
  - Exponential function:
    $\exp(x + y) = \exp(x) \cdot \exp(y)$,
    $\exp(x - y) = \exp(x)/\exp(y)$ and
    $[\exp(x)]^y = \exp(x \cdot y)$