

线性回归

Herbert002

2021 年 12 月 1 日

摘要

尽量通俗易懂，基于理工科本科数学水平的线性回归笔记.

目录

1	线性回归的通俗理解	1
2	线性回归的简要历史	1
3	一元线性回归	1
4	多元线性回归	1
5	附录	6

1 线性回归的通俗理解

线性回归的通俗理解.

2 线性回归的简要历史

高尔顿爵士关于身高体重的实验.

3 一元线性回归

我们先考虑一元线性回归.

4 多元线性回归

多元线性回归是指回归变量个数大于等于 2 的线性回归.
回归平方和服从卡方分布的证明

$$\begin{aligned}
SSR &= \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \\
&= [\hat{\mathbf{y}} - \mathbf{1}\bar{y}]^T [\hat{\mathbf{y}} - \mathbf{1}\bar{y}] \\
&= [\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} - \mathbf{1}(\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T \mathbf{y}]^T [\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} - \mathbf{1}(\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T \mathbf{y}] \\
&= \mathbf{y}^T [\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T - \mathbf{1}(\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T]^T [\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T - \mathbf{1}(\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T] \mathbf{y}
\end{aligned} \tag{1}$$

上式中, $\mathbf{1} \in \mathbb{R}^n$, 是一个元素均为 1 的列向量. 响应变量均值可以表示为 $\bar{y} = (\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T \mathbf{y}$. 接下来证明 $\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T - \mathbf{1}(\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T$ 是对称矩阵.

$$\begin{aligned}
[\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T - \mathbf{1}(\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T]^T &= [\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T]^T - [\mathbf{1}(\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T]^T \\
&= \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T - \mathbf{1}(\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T
\end{aligned} \tag{2}$$

得证.

证明 $\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T - \mathbf{1}(\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T$ 是幂等的.

如果说矩阵 A 是幂等的, 则有 $A = AA = \dots = A^n$

证明 $\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T - \mathbf{1}(\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T$ 的幂等性, 需要先证明如下结论.

$$\begin{aligned}
\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{1} &= \mathbf{1} \\
\mathbf{1}^T \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T &= \mathbf{1}^T
\end{aligned} \tag{3}$$

$\mathbf{X} \in \mathbb{R}^{n \times (k+1)}$ 可以分解为 $\mathbf{X} = [\mathbf{1} \ \mathbf{X}_R]$, 其中 $\mathbf{X}_R \in \mathbb{R}^{n \times k}$, 则有

$$\mathbf{X}^T \mathbf{X} = \begin{bmatrix} \mathbf{1}^T \\ \mathbf{X}_R^T \end{bmatrix} \begin{bmatrix} \mathbf{1} & \mathbf{X}_R \end{bmatrix} = \begin{bmatrix} \mathbf{1}^T \mathbf{1} & \mathbf{1}^T \mathbf{X}_R \\ \mathbf{X}_R^T \mathbf{1} & \mathbf{X}_R^T \mathbf{X}_R \end{bmatrix} \tag{4}$$

显然, $\mathbf{X}^T \mathbf{X}$ 是对称矩阵, 其逆矩阵也是对称矩阵, 假设其为

$$(\mathbf{X}^T \mathbf{X})^{-1} = \begin{bmatrix} a & \mathbf{b}^T \\ \mathbf{b} & \mathbf{C} \end{bmatrix} \tag{5}$$

其中, $a \in \mathbb{R}$ 是个标量, $\mathbf{b} \in \mathbb{R}^k$, $\mathbf{C} \in \mathbb{R}^{k \times k}$, 则有

$$\begin{aligned}
\mathbf{X}^T \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} &= \begin{bmatrix} \mathbf{1}^T \mathbf{1} & \mathbf{1}^T \mathbf{X}_R \\ \mathbf{X}_R^T \mathbf{1} & \mathbf{X}_R^T \mathbf{X}_R \end{bmatrix} \begin{bmatrix} a & \mathbf{b}^T \\ \mathbf{b} & \mathbf{C} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{1}^T \mathbf{1} a + \mathbf{1}^T \mathbf{X}_R \mathbf{b} & \mathbf{1}^T \mathbf{1} \mathbf{b}^T + \mathbf{1}^T \mathbf{X}_R \mathbf{C} \\ \mathbf{X}_R^T \mathbf{1} a + \mathbf{X}_R^T \mathbf{X}_R \mathbf{b} & \mathbf{X}_R^T \mathbf{1} \mathbf{b}^T + \mathbf{X}_R^T \mathbf{X}_R \mathbf{C} \end{bmatrix} \\
&= \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \mathbf{I} \end{bmatrix}
\end{aligned} \tag{6}$$

上式中 $\mathbf{0} \in \mathbb{R}^k$, $\mathbf{I} \in \mathbb{R}^{k \times k}$ 是单位矩阵. 观察最后一个等号前后矩阵的第一行, 显然有

$$\begin{aligned}
\mathbf{1}^T \mathbf{1} a + \mathbf{1}^T \mathbf{X}_R \mathbf{b} &= 1 \\
\mathbf{1}^T \mathbf{1} \mathbf{b}^T + \mathbf{1}^T \mathbf{X}_R \mathbf{C} &= \mathbf{0}^T
\end{aligned} \tag{7}$$

再看 $\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$, 如下

$$\begin{aligned}
\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{1} &= \begin{bmatrix} \mathbf{1} & \mathbf{X}_R \end{bmatrix} \begin{bmatrix} a & \mathbf{b}^T \\ \mathbf{b} & \mathbf{C} \end{bmatrix} \begin{bmatrix} \mathbf{1}^T \\ \mathbf{X}_R^T \end{bmatrix} \mathbf{1} \\
&= \begin{bmatrix} a\mathbf{1} + \mathbf{X}_R \mathbf{b} & \mathbf{1}\mathbf{b}^T + \mathbf{X}_R \mathbf{C} \end{bmatrix} \begin{bmatrix} \mathbf{1}^T \\ \mathbf{X}_R^T \end{bmatrix} \mathbf{1} \\
&= \begin{bmatrix} a\mathbf{1}\mathbf{1}^T + \mathbf{X}_R \mathbf{b}\mathbf{1}^T + \mathbf{1}\mathbf{b}^T \mathbf{X}_R^T + \mathbf{X}_R \mathbf{C} \mathbf{X}_R^T \end{bmatrix} \mathbf{1} \\
&= a\mathbf{1}\mathbf{1}^T \mathbf{1} + \mathbf{X}_R \mathbf{b}\mathbf{1}^T \mathbf{1} + \mathbf{1}\mathbf{b}^T \mathbf{X}_R^T \mathbf{1} + \mathbf{X}_R \mathbf{C} \mathbf{X}_R^T \mathbf{1}
\end{aligned} \tag{8}$$

观察上式最后一个等号的右边，第一项中 $\mathbf{1}^T \mathbf{1} \in \mathbb{R}$ ，第三项中 $\mathbf{b}^T \mathbf{X}_R^T \mathbf{1}$ 都是标量，标量的转置是其本身，再结合 (7) 式，第一项和第三项的和

$$\begin{aligned}
a\mathbf{1}\mathbf{1}^T \mathbf{1} + \mathbf{1}\mathbf{b}^T \mathbf{X}_R^T \mathbf{1} &= a(\mathbf{1}^T \mathbf{1})\mathbf{1} + (\mathbf{b}^T \mathbf{X}_R^T \mathbf{1})\mathbf{1} \\
&= a(\mathbf{1}^T \mathbf{1})\mathbf{1} + (\mathbf{1}^T \mathbf{X}_R \mathbf{b})\mathbf{1} \\
&= [a(\mathbf{1}^T \mathbf{1}) + (\mathbf{1}^T \mathbf{X}_R \mathbf{b})]\mathbf{1} \\
&= \mathbf{1}
\end{aligned} \tag{9}$$

结合 (7) 式，再看第二项和第四项的和

$$\begin{aligned}
\mathbf{X}_R \mathbf{b}\mathbf{1}^T \mathbf{1} + \mathbf{X}_R \mathbf{C} \mathbf{X}_R^T \mathbf{1} &= [(\mathbf{X}_R \mathbf{b}\mathbf{1}^T \mathbf{1} + \mathbf{X}_R \mathbf{C} \mathbf{X}_R^T \mathbf{1})^T]^T \\
&= [\mathbf{1}^T \mathbf{1}\mathbf{b}^T \mathbf{X}_R^T + \mathbf{1}^T \mathbf{X}_R \mathbf{C}^T \mathbf{X}_R^T]^T \\
&= [(\mathbf{1}^T \mathbf{1}\mathbf{b}^T + \mathbf{1}^T \mathbf{X}_R \mathbf{C}^T) \mathbf{X}_R^T]^T \\
&= [\mathbf{0}^T \mathbf{X}_R^T]^T \\
&= \mathbf{0}
\end{aligned} \tag{10}$$

回看 (8) 式，可得 $\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{1} = \mathbf{1}$ ，同理可得 $\mathbf{1}^T \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T = \mathbf{1}^T$ 。推导过程中，需要注意矩阵和向量的维数。

接下来就有

$$\begin{aligned}
&[\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T - \mathbf{1}(\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T]^2 \\
&= [\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T - \mathbf{1}(\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T][\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T - \mathbf{1}(\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T] \\
&= \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T - \mathbf{1}(\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \\
&\quad - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{1}(\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T + \mathbf{1}(\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T \mathbf{1}(\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T \\
&= \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T - \mathbf{1}(\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T - \mathbf{1}(\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T + \mathbf{1}(\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T \\
&= \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T - \mathbf{1}(\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T
\end{aligned} \tag{11}$$

结合 (2) 式的结论，(1) 式可继续推导为

$$\begin{aligned}
SSR &= \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \\
&= \mathbf{y}^T [\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T - \mathbf{1}(\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T]^T [\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} - \mathbf{1}(\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T \mathbf{y}] \\
&= \mathbf{y}^T [\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T - \mathbf{1}(\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T] [\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T - \mathbf{1}(\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T] \mathbf{y} \\
&= \mathbf{y}^T [\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T - \mathbf{1}(\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T] \mathbf{y}
\end{aligned} \tag{12}$$

至此，上面已经证明了 $\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T - \mathbf{1}(\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T$ 是一个幂等的实对称矩阵。接下来，将其简单表示为 \mathbf{Q} ，则有

$$SSR = \mathbf{y}^T \mathbf{Q} \mathbf{y} \quad (13)$$

因为 $\mathbf{Q} \in \mathbb{R}^{n \times n}$ 是实对称矩阵，所以其可以正交对角化，即存在正交矩阵 $\mathbf{P} \in \mathbb{R}^{n \times n}$ ，使得

$$\mathbf{P}^{-1} \mathbf{Q} \mathbf{P} = \mathbf{P}^T \mathbf{Q} \mathbf{P} = \mathbf{\Lambda} \quad (14)$$

其中， $\mathbf{\Lambda} \in \mathbb{R}^{n \times n}$ 是实对角矩阵，对角线上的元素是 \mathbf{Q} 的 n 个特征值。所以有

$$\begin{aligned} SSR &= \mathbf{y}^T \mathbf{Q} \mathbf{y} \\ &= \mathbf{y}^T \mathbf{P} \mathbf{\Lambda} \mathbf{P}^T \mathbf{y} \\ &= (\mathbf{P}^T \mathbf{y})^T \mathbf{\Lambda} (\mathbf{P}^T \mathbf{y}) \end{aligned} \quad (15)$$

又因为 \mathbf{Q} 是等幂矩阵，所以有

$$\begin{aligned} \mathbf{\Lambda}^2 &= \mathbf{P}^T \mathbf{Q} \mathbf{P} \mathbf{P}^T \mathbf{Q} \mathbf{P} \\ &= \mathbf{P}^T \mathbf{Q}^2 \mathbf{P} \\ &= \mathbf{P}^T \mathbf{Q} \mathbf{P} \\ &= \mathbf{\Lambda} \end{aligned} \quad (16)$$

可见 $\mathbf{\Lambda}$ 也是等幂的。如果一个对角矩阵是等幂的，其对角元素要么是 1，要么是 0（可自行证明）。

用 p_{ij} 表示 \mathbf{P} 第 i 行第 j 列的元素。 $\mathbf{P}^T \mathbf{y} \in \mathbb{R}^n$ 的第 i 个元素可以表示为 $\sum_{j=1}^n p_{ji} y_j$ ，带入 (15) 式，可得

$$\begin{aligned} SSR &= (\mathbf{P}^T \mathbf{y})^T \mathbf{\Lambda} (\mathbf{P}^T \mathbf{y}) \\ &= \begin{bmatrix} \sum_{j=1}^n p_{j1} y_j & \sum_{j=1}^n p_{j2} y_j & \cdots & \sum_{j=1}^n p_{jn} y_j \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \begin{bmatrix} \sum_{j=1}^n p_{j1} y_j \\ \sum_{j=1}^n p_{j2} y_j \\ \vdots \\ \sum_{j=1}^n p_{jn} y_j \end{bmatrix} \\ &= \sum_{i=1}^n \lambda_i \left(\sum_{j=1}^n p_{ji} y_j \right)^2 \\ &= \sum_{i=1}^m \left(\sum_{j=1}^n p_{ji} y_j \right)^2 \end{aligned} \quad (17)$$

上式中， m 是对角矩阵 $\mathbf{\Lambda}$ 对角元素为 1 的个数。

当 $\boldsymbol{\beta} = \mathbf{0}$ 时， $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} = \boldsymbol{\epsilon}$ ，则有

$$SSR = \sum_{i=1}^m \left(\sum_{j=1}^n p_{ji} y_j \right)^2 = \sum_{i=1}^m \left(\sum_{j=1}^n p_{ji} \epsilon_j \right)^2 \quad (18)$$

则有

$$\begin{aligned}
& \epsilon_j \sim \mathcal{N}(0, \sigma^2) \\
& \Downarrow \\
& p_{ji}\epsilon_j \sim \mathcal{N}(0, p_{ji}^2 \sigma^2) \\
& \Downarrow \\
& \sum_{j=1}^n p_{ji}\epsilon_j \sim \mathcal{N}(0, \sum_{i=1}^n p_{ji}^2 \sigma^2) \\
& \Downarrow \\
& \sum_{j=1}^n p_{ji}\epsilon_j \sim \mathcal{N}(0, \sigma^2 \sum_{i=1}^n p_{ji}^2) \\
& \Downarrow \left(\sum_{i=1}^n p_{ji}^2 = 1 \right) \\
& \sum_{j=1}^n p_{ji}\epsilon_j \sim \mathcal{N}(0, \sigma^2) \\
& \Downarrow \\
& \frac{1}{\sigma} \sum_{j=1}^n p_{ji}\epsilon_j \sim \mathcal{N}(0, 1) \\
& \Downarrow \\
& \sum_{i=1}^m \left(\frac{1}{\sigma} \sum_{j=1}^n p_{ji}\epsilon_j \right)^2 \sim \chi^2(m) \\
& \Downarrow \\
& \frac{1}{\sigma^2} SSR \sim \chi^2(m)
\end{aligned} \tag{19}$$

上面倒数第二个推导 (\Downarrow) 需要证明 $\frac{1}{\sigma} \sum_{j=1}^n p_{ji}\epsilon_j$ 和 $\frac{1}{\sigma} \sum_{j=1}^n p_{jl}\epsilon_j$ 之间的独立性, 如下

$$\begin{aligned}
Cov\left(\frac{1}{\sigma} \sum_{j=1}^n p_{ji}\epsilon_j, \frac{1}{\sigma} \sum_{j=1}^n p_{jl}\epsilon_j\right) &= \frac{1}{\sigma^2} Cov\left(\sum_{j=1}^n p_{ji}\epsilon_j, \sum_{j=1}^n p_{jl}\epsilon_j\right) \\
&= \frac{1}{\sigma^2} \sum_{j=1}^n \sum_{h=1}^n Cov(p_{ji}\epsilon_j, p_{hl}\epsilon_h) \\
&= \frac{1}{\sigma^2} \sum_{j=1}^n \sum_{h=1}^n p_{ji}p_{hl}Cov(\epsilon_j, \epsilon_h) \\
&= \frac{1}{\sigma^2} \sum_{j=1}^n \sum_{h=1}^n p_{ji}p_{hl} \cdot 0 \\
&= 0
\end{aligned} \tag{20}$$

上式中, 因为 ϵ_i 之间相互独立, 所以 $Cov(\epsilon_i, \epsilon_j) = 0$. 协方差为 0 的正态分布, 相互独立. 独立性得证.

最后, 来看看上面这个卡方分布的自由度 m , 到底是多少. 上面已经说过, 是对角矩阵 \mathbf{A} 的对角元素要么是 1, 要么是 0, m 是 1 的个数, 也就是说, m 也是对角矩阵 \mathbf{A} 的迹, 所以

$$\begin{aligned}
m &= \text{trace}(\mathbf{A}) \\
&= \text{trace}(\mathbf{P}^T \mathbf{Q} \mathbf{P}) \\
&= \text{trace}(\mathbf{P} \mathbf{P}^T \mathbf{Q}) \\
&= \text{trace}(\mathbf{Q}) \\
&= \text{trace}[\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T - \mathbf{1}(\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T] \\
&= \text{trace}[\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T] - \text{trace}[\mathbf{1}(\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T] \\
&= \text{trace}[\mathbf{X}^T \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1}] - \text{trace}[\mathbf{1}^T \mathbf{1}(\mathbf{1}^T \mathbf{1})^{-1}] \\
&= \text{trace}(\mathbf{I}_{k+1}) - \text{trace}(\mathbf{I}_1) \\
&= k + 1 - 1 \\
&= k
\end{aligned} \tag{21}$$

上式中的 k 就是回归模型中回归变量的个数. 推导过程中使用到了矩阵迹的性质。

到此，我们已经基本完整地证明了回归平方和 $\frac{1}{\sigma^2} SSR$ 服从自由度为 k 的卡方分布，用数学表达式表示为

$$\frac{1}{\sigma^2} SSR \sim \chi^2(k) \tag{22}$$

5 附录

定理一

实对称矩阵 $\mathbf{A} \in \mathbb{R}^{n \times n}$ 必定正交相似于实对角矩阵 $\mathbf{\Lambda}$ ，即存在正交矩阵 \mathbf{P} ，使得 $\mathbf{P}^{-1} \mathbf{A} \mathbf{P} = \mathbf{P}^T \mathbf{A} \mathbf{P} = \mathbf{\Lambda}$ ，其中 $\mathbf{\Lambda}$ 对角线上的元素是 \mathbf{A} 的 n 个特征值。

证明可请参考线性代数的教材。

定理二

矩阵迹的运算有如下性质：

$$\begin{aligned}
\text{trace}(\mathbf{A} \mathbf{B}) &= \text{trace}(\mathbf{B} \mathbf{A}) \\
\text{trace}(\mathbf{A} + \mathbf{B}) &= \text{trace}(\mathbf{A}) + \text{trace}(\mathbf{B}) \\
\text{trace}(c \mathbf{A}) &= c \cdot \text{trace}(\mathbf{A})
\end{aligned} \tag{23}$$