

نگاهی به معیارهای میانی و پراکندگی

خسرو فضلی*

جواد بهبودیان†

چکیده

زیرا اگر آن را به جای هر x_i بگذاریم مجموع داده‌ها تغییر نمی‌کند. به سخنی دیگر مجموع داده‌ها پایا می‌ماند. از این‌رو \bar{x} را یک معیار میانی با «پایایی مجموع» می‌نامیم. این ویژگی در جدول زیر نمایان است.

$$\begin{array}{c|cc} & \text{داده‌ها} \\ \hline x_1, x_2, \dots, x_n & \sum_{i=1}^n x_i \\ \bar{x} & n\bar{x} \end{array} \quad \begin{array}{l} | \\ \text{میانگین حسابی} \end{array}$$

اینک با فرض مثبت بودن x_i ‌ها یک معیار میانی با «پایایی حاصلضرب» می‌سازیم.

این معیار را با G نشان می‌دهیم. اگر G را به جای هر x_i بگذاریم باید حاصلضرب داده‌ها تغییر نکند، یعنی

$$\prod_{i=1}^n x_i = \prod_{i=1}^n G = G^n$$

بنابراین میانگین هندسی به صورت زیر به دست می‌آید:

$$G = \sqrt[n]{x_1 x_2 \dots x_n}.$$

این دو مثال نشان می‌دهند که ویژگی مطلوب را می‌توان به صورت یک تابع n متغیری متقارن از داده‌ها نشان داد. برای مثال اول این تابع به صورت

$$f(x_1, x_2, \dots, x_n) = x_1 + x_2 + \dots + x_n$$

و برای مثال دوم به صورت

$$f(x_1, x_2, \dots, x_n) = x_1 x_2 \dots x_n$$

است. به طور کلی، یا تابع n متغیری را متقارن می‌گویند، هرگاه برای هر

۱ پیشگفتار

داده‌های آماری را معمولاً در عددی به نام معیار میانی، یا مرکزی، مانند میانگین، میانی یا نما، خلاصه می‌کنند. این عدد در میان داده‌ها جا دارد و در حکم نماینده آنهاست. افزون بر این، عددی دیگر به نام معیار پراکندگی، مانند برد یا واریانس، را به کار می‌برند که میزان تغییر میان داده‌ها یا چگونگی پراکندگی آنها را نسبت به یک معیار میانی بیان می‌دارد. معیار میانی و معیار پراکندگی را می‌توان به راه‌های گوناگون با ویژگی‌های مشخص تعریف کرد.

در این مقاله می‌خواهیم با یک نگاه بر پایه یک تابع چند متغیری متقارن، تمام معیارهای میانی را بیان داریم. همین کار را در مورد معیارهای پراکندگی بر پایه دو تابع متقارن انجام می‌دهیم. برای روشن شدن موضوع مثالهای گوناگون ارائه می‌دهیم.

۲ معیارهای میانی در یک نگاه

معیار میانی عددی است که داده‌های x_1, x_2, \dots, x_n را بر پایه یک ویژگی مناسب خلاصه نماید. مثلاً \bar{x} (میانگین حسابی) یک معیار میانی است،

* خسرو فضلی گروه ریاضی دانشگاه کردستان † دکتر جواد بهبودیان بخش آمار، دانشگاه شیراز

$$f(x_1, x_2, \dots, x_n) = x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}$$

و در نتیجه داریم:

$$x^* = \frac{1}{2}[x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}]$$

در جدول زیر چند معیار میانی داده شده‌اند.

جدول معیارهای میانی

تابع متقارن	داده‌ها	x^*	نام معیار
$x_1 + x_2 + \dots + x_n$	حقیقی	$\frac{1}{n} \sum_{i=1}^n x_i$	میانگین حسابی
$x_1 x_2 \dots x_n$	متبت	$\sqrt[n]{x_1 x_2 \dots x_n}$	میانگین هندسی
$\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}$	متبت	$\sum_{i=1}^n \frac{1}{x_i}$	میانگین توانی
$x_1^r + x_2^r + \dots + x_n^r$	متبت	$\sqrt[n]{\frac{1}{n} \sum_{i=1}^n x_i^r}$	میانگین ریشه‌ای رتبه r
$e^{x_1} + e^{x_2} + \dots + e^{x_n}$	حقیقی	$\log \bar{x}$	میانگین نمایی
$x_{(\frac{n+1}{2})}$	حقیقی n فرد	$x_{(\frac{n+1}{2})}$	میانه
$\sqrt{\sum_{i \neq j} x_i x_j}$	متبت	$\sqrt{\frac{\sum_{i \neq j} x_i x_j}{n(n-1)}}$	میانگین حاصلضربی

لازم است بیفزاییم که در جدول بالا، پنج معیار نخست حالتهای ویژه‌ای میانگین زیر به نام میانگین تعیین یافته‌اند:

$$g(x_1, x_2, \dots, x_n) = h^{-1}\left(\frac{1}{n} \sum_{i=1}^n h(x_i)\right)$$

در این رابطه تابع h یک تابع بیوسته یکواری اکید با وارون h^{-1} و تابع g متقارن است. برای پنج معیار بالا تابع h به ترتیب عبارت است از x و $\frac{1}{x}$ و x^r و e^x . ولی دو معیار آخر جدول را نمی‌توان از میانگین تعیین یافته به دست آورد. بنابراین تعریف کلی بالا، معیارهای بیشتری را در بر می‌گیرد.

۳ معیارهای پراکندگی در یک نگاه

معیار پراکندگی عددی است که میزان تفاوت میان داده‌های x_1, x_2, \dots, x_n ، $n > 1$ ، را نسبت به یک معیار میانی یا نسبت به یکدیگر بیان می‌دارد. برای تعیین چنین معیاری، نخست میزان دوری دو کیت a و b را به وسیله یک تابع فاصله‌ای $d(a, b)$ ، که نامنفی و متقارن است، مانند $|a - b|$ یا $|a - b|^p$ ، $p > 0$ ، می‌سنجیم. اینک دو حالت را بررسی می‌کنیم:

الف - پراکندگی نسبت به یک معیار میانی

فرض کنید x^* یک معیار میانی با پایابی f برای داده‌ها باشد. میزان دوری داده‌ها نسبت به x^* بر پایه تابع فاصله‌ای $d(a, b)$ عبارت‌اند از

$$y_1 = d(x_1, x^*), y_2 = d(x_2, x^*), \dots, y_n = d(x_n, x^*).$$

جایگشت دلخواه x_1, x_2, \dots, x_n از $x_{in}, x_{in+1}, \dots, x_{in+n-1}$ داشته باشیم:

$$f(x_1, x_2, \dots, x_n) = f(x_{in}, x_{in+1}, \dots, x_{in+n-1})$$

متلاً برای x_1, x_2, \dots, x_n تابع سه متغیری $x_1^*, x_2^*, \dots, x_n^*$ متقارن ولی تابع سه متغیری $x_1^* + x_2^* + \dots + x_n^*$ نامتقارن است. ویژگی مطلوب را با تابع متقارن بیان می‌دارند، زیرا معیار میانی نباید با جایگشت داده‌ها تغییر کند.

از آنجه در بالاگفتیم، می‌توان تعریف کلی زیر را برای معیار میانی ارائه داد.

تعریف معیار میانی. فرض کنید $R^n \rightarrow A : f$ ، که در آن $A \subset R^n$ یک تابع n متغیری متقارن باشد. عدد x^* را یک معیار میانی با «پایابی f » می‌نامند، هرگاه برای هر n تایی (x_1, x_2, \dots, x_n) در A داشته باشیم:

$$f(x_1, x_2, \dots, x_n) = f(x_1^*, x_2^*, \dots, x_n^*)$$

بدیهی است که n تایی $(x_1^*, x_2^*, \dots, x_n^*)$ هم باید در A باشد. اینک به ذکر دو مثال دیگر می‌پردازیم.

مثال ۱. معیار میانی را برای داده‌های نامنفی با پایابی زیر بیابید:

$$f(x_1, x_2, \dots, x_n) = \sum_{i \neq j} x_i x_j$$

حل. تابع داده شده متقارن است. از حل معادله

$$\sum x_i x_j = \sum_{i \neq j} x_i^* x_j^* = n(m-1)x^*$$

بر حسب x^* داریم:

$$x^* = \sqrt{\frac{\sum_{i \neq j} x_i x_j}{n(n-1)}}$$

این معیار را میانگین حاصلضربی می‌نامیم.

مثال ۲. داده‌ها را به صورت $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ مرتب می‌کنیم. معیار میانی را با پایابی زیر، در صورتی که n فرد باشد، بیابید:

$$f(x_1, x_2, \dots, x_n) = x_{(\frac{n+1}{2})}$$

حل. تابع داده شده متقارن است. از حل معادله

$$f(x_1, x_2, \dots, x_n) = f(x^*, x^*, \dots, x^*)$$

بر حسب x^* داریم:

$$x^* = x_{(\frac{n+1}{2})}$$

این معیار میانه داده‌های است. اگر n زوج باشد، تابع را به صورت زیر تعریف می‌کنیم:

ب - پراکندگی داده‌ها نسبت به یکدیگر

اگر دوری داده‌ها را نسبت به یکدیگر بر پایه تابع فاصله‌ای $d(a, b)$ در نظر بگیریم، $(1 - n/n)$ عدد به صورت $d(x_i, x_j)$, $i \neq j$, به دست می‌آید.

حال بر پایه متقارن (t_1, t_2, \dots, t_n) , $g(t_1, t_2, \dots, t_n) = g(t_1, t_2, \dots, t_n)$, معیار میانی این اعداد را می‌باییم تا v^* یعنی معیار پراکندگی داده‌ها نسبت به یکدیگر به دست آید.

مثال ۵. بر پایه $d(a, b) = |a - b|$ و $d(a, b) = (a - b)^r$, $g(t_1, t_2, \dots, t_n) = \sum_{i=1}^{n(n-1)} |t_i - t_{i+1}|^r$ یک معیار پراکندگی برای x_1, x_2, \dots, x_n نسبت به یکدیگر بیاید.

حل - به آسانی داریم

$$v^* = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i} (x_i - x_j)^r$$

مثال ۶. بر پایه $d(a, b) = |a - b|$ و $d(a, b) = (a - b)^r$, $g(t_1, t_2, \dots, t_n) = \max_{i \in \{1, 2, \dots, n\}} |t_i - t_{i+1}|^r$ یک معیار پراکندگی برای x_1, x_2, \dots, x_n نسبت به یکدیگر بیاید.

حل - v^* را باید به گونه‌ای بگیریم تا داشته باشیم:

$$\begin{aligned} g(|x_1 - x_2|, |x_1 - x_3|, \dots, |x_{n-1} - x_n|) \\ = g(v^*, v^*, \dots, v^*) \end{aligned}$$

اگر $x_{(1)}$ و $x_{(n)}$ به ترتیب کوچکترین و بزرگترین مقادیر داده باشند، از حل معادله بالا بر حسب v^* , برد داده‌ها یعنی

$$v^* = x_{(n)} - x_{(1)}$$

به دست می‌آید.

حال معیار میانی این دورها را با پایانی (y_1, y_2, \dots, y_n) $g(y_1, y_2, \dots, y_n)$ می‌باییم و آن را v^* نامیم. بنابراین با داشتن معیار میانی x^* , یک معیار پراکندگی نسبت به x^* بر پایه تابع فاصله‌ای d و تابع متقارن g به دست می‌آید.

مثال ۳. فرض کنید $\bar{x} = x^*$ معیار میانی برای x_1, x_2, \dots, x_n باشد. بر پایه تابع فاصله‌ای $d(a, b) = |a - b|$ و تابع متقارن $g(t_1, t_2, \dots, t_n) = \sum_{i=1}^n |t_i - \bar{x}|^r$ یک معیار پراکندگی نسبت به \bar{x} به دست آورید.

حل - باید v^* را به گونه‌ای بیابیم تا داشته باشیم:

$$g(d(x_1, \bar{x}), d(x_2, \bar{x}), \dots, d(x_n, \bar{x})) = g(v^*, v^*, \dots, v^*)$$

$$\sum_{i=1}^n (x_i - \bar{x})^r = \sum_{i=1}^n v^* = nv^*$$

بنابراین واریانس داده‌ها به صورت زیر به دست می‌آید:

$$v^* = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^r$$

مثال ۴. فرض کنید $x^* = x_{(\frac{n+1}{2})} = m$ معیار میانی برای x_1, x_2, \dots, x_n با فرض فرد بودن n باشد. بر پایه $d(a, b) = |a - b|$ و $g(t_1, \dots, t_n) = \sum_{i=1}^n |t_i - m|^r$, یک معیار پراکندگی نسبت به m بیاید.

حل - به آسانی به دست می‌آوریم:

$$v^* = \frac{1}{n} \sum_{i=1}^n |x_i - m|^r.$$

مراجع

- [2] Berger, R. L. and Casella, G. (1992), "Deriving Generalized Means as least Squares and Maximum Likelihood Estimates", *The American Statistician*, Vol. 46, No. 4, 279-282.

- [۱] جواد بهبودیان، آمار و احتمال مقدماتی، چاپ نهم، ۱۳۷۵، آستان قدس رضوی.