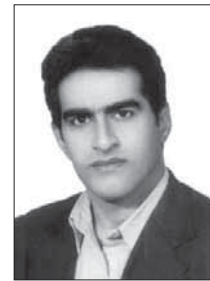


دسته‌بندی داده‌های پیوسته در آمار توصیفی

(درس آمار و مدل سازی دوره دوم متوسطه)

مقدمه

یکی از بخش‌های علم آمار، آمار توصیفی است که کتاب «آمار و مدل سازی» به آن پرداخته است. آمار توصیفی شامل روش‌های جمع‌آوری و خلاصه کردن داده‌هاست و یکی از راه‌های خلاصه کردن داده‌های متغیر تصادفی پیوسته، دسته‌بندی (طبقه‌بندی) آن‌هاست. دسته‌بندی داده‌ها در کتاب‌های آمار به روش‌هایی متفاوت مطرح شده است، ولی هدف همه آن‌ها کم کردن حجم داده‌ها برای راحت توصیف کردن موضوع یا موضوعات مورد مطالعه است. در این مقاله سعی شده روشی را که کتاب درسی آمار و مدل سازی برای دسته‌بندی داده‌ها مطرح کرده است، با روش‌های کتاب‌های دیگر آماری تلفیق کنیم و روشی ساده و راحت برای دانش‌آموزان و دانشجویان مطرح سازیم.



رضا زینی‌وند
دبیر دبیرستان‌های
استان ایلام - دره‌شهر

کلیدواژه‌ها: داده‌های آماری، دسته‌بندی

۲. مشخص کردن تعداد دسته‌ها

تعداد دسته‌ها را می‌توان به‌طور اختیاری با توجه به تعداد داده‌ها بین ۵ تا ۲۰ دسته تعیین کرد. همچنین می‌توان تعداد دسته‌ها را نیز با استفاده از فرمول‌هایی به‌صورت زیر محاسبه کرد (این دستورها پیشنهادی بوده و دقیق نمی‌باشند و تنها می‌توانند نقطه شروعی مناسب را برای دسته‌بندی به‌دست دهند):

الف. محاسبه تعداد دسته‌ها با فرمول $k = [\sqrt{n}] + 1$

که در آن k تعداد دسته‌ها، n تعداد داده‌ها و نماد $[\]$ جزء صحیح است. در مثال ۱ تعداد دسته‌ها به‌صورت زیر محاسبه می‌شود:

$$k = [\sqrt{30}] + 1 = [5.477...] + 1 = 6$$

مراحل دسته‌بندی داده‌ها

مراحل دسته‌بندی داده‌ها را با یک مثال مطرح می‌کنیم.

مثال ۱. داده‌های زیر مربوط به نمرات پایانی اول درس آمار و مدل سازی یک کلاس ۳۰ نفره است:

۱۳/۷۵	۷/۵۰	۱۵/۲۵	۱۶	۸/۵۰	۱۴/۲۵	۱۶	۱۲	۱۲/۷۵	۱۴/۵۰
۱۲/۵۰	۶/۲۵	۱۲/۵۰	۱۹	۱۷/۲۵	۱۳/۷۵	۱۸	۱۴	۱۵/۵۰	۱۸
۱۴	۷/۵۰	۱۲	۱۶/۵۰	۱۷	۱۳/۲۵	۱۶	۱۵/۵۰	۱۴	۱۵

۱. محاسبه دامنه تغییرات داده‌ها

حداکثر اختلاف داده‌ها را دامنه تغییرات گوئیم که به‌صورت زیر محاسبه می‌شود:

$$R = b - a = 19 - 6/25 = 12/25$$

که در آن b بزرگ‌ترین و a کوچک‌ترین داده است.



ب. محاسبه تعداد دسته‌ها (k) به طوری که کوچک‌ترین عدد صحیح باشد که در رابطه $n \leq 2^k$ صدق کند. در مثال ۱ این مقدار برای $n=30$ برابر ۵ است، چون داریم: $2^4 < 30 < 2^5$. در مثال ۱ تعداد دسته‌ها را ۵ در نظر می‌گیریم. هرچند برای تعیین تعداد دسته هیچ‌گونه روش دقیق ریاضی وجود ندارد، ولی فرمول‌های فوق می‌توانند به مشخص کردن تعداد دسته‌ها کمک کنند.

$$c = \frac{12/75}{5} = 2/55 \approx 2/56$$

۴. تعیین کران‌های دسته‌ها

برای تعیین کران‌های دسته‌ها، ابتدا کران پایین دسته اول را به این صورت محاسبه می‌کنیم که مقدار طول دسته به دست آمده در بخش قبلی را در تعداد دسته‌ها ضرب می‌کنیم و آن را R' می‌نامیم. با

۳. تعیین طول دسته‌ها

طول دسته‌ها همان طوری که در کتاب درسی آمار و مدل‌سازی آمده است، از تفاضل کران‌های پایین (کران‌های بالا) یا تفاضل نماینده (مرکز) متوالی دسته‌ها به دست می‌آید.

محاسبه نصف تفاضل $R'-R$ مقداری به دست می آید که در محاسبه کران‌ها از آن استفاده می‌کنیم. در مثال ۱ داریم:

$$\frac{R'-R}{2} = \frac{2/56 \times 5 - 12/55}{2} = 0.025$$

حال مقدار به دست آمده را از کوچک‌ترین داده کم می‌کنیم تا کران پایین دسته اول مشخص شود. در مثال ۱ داریم:

کران پایین دسته اول:

$$6/25 - 0.025 = 6/225$$

حال با اضافه کردن مقدار طول دسته (c) به کران پایین دسته اول، کران پایین دسته دوم مشخص می‌شود. به همین ترتیب کران پایین دسته‌های بعدی نیز مشخص می‌شود. در مثال ۱ داریم:

کران پایین دسته دوم:

$$6/225 + 2/56 = 8/785$$

کران پایین دسته سوم:

$$8/785 + 2/56 = 11/345$$

کران پایین دسته چهارم:

$$11/345 + 2/56 = 13/905$$

کران پایین دسته پنجم:

$$13/905 + 2/56 = 16/465$$

بعد از تعیین کران‌های پایین دسته‌ها، برای تعیین کران بالای دسته‌ها می‌توان با اضافه کردن طول دسته به کران‌های پایین، کران‌های بالا را مشخص کرد. در مثال ۱ کران‌ها به صورت زیر محاسبه می‌شوند:

کران بالای دسته اول =

کران پایین دسته اول + طول دسته:

$$6/225 + 2/56 = 8/785$$

کران بالای دسته دوم =

کران پایین دسته دوم + طول دسته:

$$8/785 + 2/56 = 11/345$$

کران بالای دسته سوم =

کران پایین دسته سوم + طول دسته:

$$11/345 + 2/56 = 13/905$$

کران بالای دسته چهارم =

کران پایین دسته چهارم + طول دسته:

$$13/905 + 2/56 = 16/465$$

کران بالای دسته پنجم =

کران پایین دسته پنجم + طول دسته:

$$16/465 + 2/56 = 19/025$$

در آخر با مقایسه کران بالای دسته آخر و بزرگ‌ترین داده، صحت دسته‌بندی داده‌ها را بررسی می‌کنیم. در این مقایسه نباید کران بالای دسته آخر کمتر از بزرگ‌ترین داده باشد. همان‌طور که معلوم است، کران بالای هر دسته با کران پایین دسته بعدی برابر است که نشان‌دهنده پیوستگی داده‌هاست.

نتیجه‌گیری

از مزایای این روش دسته‌بندی، نسبت به روش کتاب درسی آمار و مدل‌سازی، یکی این است که دقت ثبت کران‌های دسته‌ها (تعداد ارقام اعشاری) یک رقم بیشتر از داده‌هاست. این وضعیت در تعیین فراوانی دسته‌ها مشکل شمارش داده‌ای را برای دو دسته رفع می‌کند. به عبارت دیگر، برای مشخص کردن فراوانی مطلق دسته‌ها، این قرارداد که «کران‌های پایین دسته‌ها در صورت امکان برای دسته‌ها شمرده شوند و کران‌های بالای دسته‌ها به جز دسته آخر شمرده نشوند» لازم نیست. مزیت دیگر این روش آن است که بین کران پایین دسته اول و کران بالای دسته آخر به ترتیب نسبت به کوچک‌ترین داده و بزرگ‌ترین داده تعادل مقداری به وجود می‌آید.

* منابع

۱. بخشعلی‌زاده، شهرناز؛ پاشا، عین‌الله؛ رستگار، آرش (۱۳۹۰). آمار و مدل‌سازی (چاپ جدید). شرکت چاپ و نشر کتاب‌های درسی ایران. تهران.
۲. نیکوکار، مسعود و عربزاده، بهمن (۱۳۸۶). آمار و احتمالات کاربردی. انتشارات آزاده، تهران. چاپ سیزدهم.
۳. روغنی شهرکی، قهرمان (۱۳۸۸). آمار و احتمالات کاربردی. انتشارات مینای خرد. تهران. چاپ اول.
۴. آذر، عادل و مؤمنی، منصور (۱۳۸۴). آمار و کاربرد آن در مدیریت (ج ۱). انتشارات سمت. تهران. چاپ یازدهم.
۵. جمشیدی، خدیجه (۱۳۸۶). آمار و کاربرد آن در مدیریت. انتشارات رادنگار. تهران. چاپ سیزدهم.