

فصل دوم

گردآوری و دسته بندی داده ها

11

توزیع یک متغیر کیفی

جنس نما	خط نشان	فراوانی
سنگ		۲۳
شیشه		۸
آجر		۱۹

12

توزیع یک متغیر کمی ناپیوسته (گسسته)

تعداد طبقات	خط نشان	فراوانی
۱		۱۲
۲		۲۰
۳		۱۱
۴		۷

13

توزیع متغیر کمی پیوسته

- وقتی متغیری پیوسته است، مرتب کردن داده های آن کمی پیچیده تر می شود، زیرا در این حالت، مقادیر بین دو عدد طبیعی نیز لحاظ می گردد. مراحل کار به این قرار است:
1. ابتدا کمترین و بیشترین مقدار را مشخص کرده و تفاضل آن ها را محاسبه می کنیم و آن را با R نشان می دهیم.

$$R = X_{\max} - X_{\min}$$

2. برای محاسبه ارتفاع طبقه (h) تفاضل X_{\max} و X_{\min} را بر تعداد طبقات مورد نظر تقسیم می کنیم (k).

$$h = R \div k$$

14

❖ برای پیدا کردن مناسب ترین تعداد طبقات (k) می توان از این فرمول استفاده کرد: $k = 1 + 3.3 \log n$

تعداد داده ها یا حجم نمونه $n =$

3. گروه اول و آخر را طوری انتخاب می کنیم که کوچکترین مقدار (x_{min}) در وسط گروه اول و بزرگترین مقدار (x_{max}) در وسط گروه آخر قرار گیرد. به این ترتیب یک گروه به تعداد طبقات اضافه می شود. بنابراین k که تعداد طبقات دلخواه است را منهای یک می کنیم و از رابطه $h = \frac{R}{k-1}$ برای محاسبه ارتفاع طبقه استفاده می کنیم.

15

4. برای مشخص کردن حد پایین طبقه اول به شکل زیر عمل

$$x_1 = x_{min} - \frac{h}{2} \quad \text{می کنیم:}$$

5. برای محاسبه حد پایین طبقات بعدی از فرمول زیر استفاده

$$x_{k+1} = x_k + h \quad \text{می کنیم:}$$

به این ترتیب حد بالا و پایین تمام طبقات به دست می آید.

❖ در عمل، k معمولاً عددی بین ۵ و ۲۰ در نظر گرفته می شود.

16

نکاتی درباره توزیع متغیر کمی

- در مرحله سوم ممکن است برای این که کوچکترین مقدار (X_{\min}) در وسط طبقه اول قرار گیرد، با مشکل مواجه شویم. الزامی برای این وجود ندارد که X_{\min} دقیقاً وسط طبقه اول قرار گیرد. هدف آن است که اطمینان حاصل شود که X_{\min} در طبقه اول قرار گرفته باشد. به این منظور برای ساده تر کردن روش قبل به ترتیب زیر عمل می کنیم:
- مرحله اول و دوم را طبق شیوه قبل انجام می دهیم
- 3. اگر حاصل R/k عددی اعشاری بود، آن را به اولین عدد صحیح بعدی گرد می کنیم.
- 4. اگر X_{\min} عددی صحیح بود، آن را به عنوان حد پایین طبقه اول در نظر می گیریم، در غیر این صورت، اولین عدد صحیح یا اولین عدد رند را که کمتر از X_{\min} است به عنوان حد پایین طبقه اول در نظر می گیریم.

17

انواع طبقه بندی داده ها

- طبقه بندی پیوسته $[X_1, X_2), [X_2, X_3), [X_3, X_4]$
مثال: 0-10, 10-20, 20-30
- طبقه بندی گسسته $[X_1, X_2], [X_3, X_4], [X_5, X_6]$
مثال: 0-10, 11-20, 21-30

18

مفاهیم تکمیلی

- اگر طبقه $[x_1, x_2]$ را در نظر بگیرید، x_1 حد پایین و x_2 حد بالا نامیده می شوند.
- حد وسط هر طبقه (Mid Point) از رابطه زیر تعیین می گردد:

$$MP_i = \frac{\text{حد پایین طبقه } i + \text{حد پایین طبقه } (i + 1)}{2}$$

- حد وسط یا MP در حکم نماینده طبقه می باشد.
- تفاضل دو حد پایین متوالی یا دو حد بالای متوالی را ارتفاع طبقه (h) می نامند.

19

مفاهیم تکمیلی

- فراوانی مطلق (Absolute frequency)
فراوانی مطلق طبقه i را که با f_i نشان می دهیم برابر تعداد اعضای است که در طبقه i قرار گرفته اند.
- فراوانی تجمعی (Cumulative frequency)
پس از مشخص شدن فراوانی مطلق هر طبقه یا f_i اگر این فراوانی ها به طور متوالی با هم جمع شوند و در طبقه مربوطه ثبت شود، فراوانی تجمعی به دست می آید. به عبارت دیگر فراوانی تجمعی عبارتست از تعداد افراد نمونه که مقدار آن ها از حد بالای طبقه مورد نظر کمتر است. فراوانی تجمعی آخرین طبقه همیشه برابر با n خواهد بود.
$$cf_i = f_1 + f_2 + \dots + f_i$$
- فراوانی نسبی (Relative frequency): سهم هر یک از طبقات
از کل حجم نمونه می باشد.
$$rf_i = f_i/n$$

20

- مثال: معدل دانشجویان شهرسازی به نحو زیر طبقه بندی شده است، مطلوب است محاسبه حد وسط (MP)، فراوانی نسبی، فراوانی تجمعی و فراوانی تجمعی نسبی.

معدل	فراوانی مطلق
12-14	3
14-16	4
16-18	9
18-20	4

21

- تمرین: مساحت قطعات یک محله به ترتیب ذیل به دست آمده است. آن ها را در تعداد طبقات مناسب، دسته بندی کنید و فراوانی و حد وسط هر طبقه را مشخص کنید.

153, 154.6, 155, 151, 150.9, 162.1, 164.9, 163.1,
 166, 169.95, 168.8, 160, 163, 167.5, 170, 172.5,
 176, 177, 178.1, 179.7, 175, 181, 181.3, 189.8,
 185, 184.9, 190, 199.6, 192, 194, 194.9, 195,
 197.3, 201, 205, 209.5, 208, 210, 211.1, 212,
 215, 217.5

22

- تمرین: مجموعه داده های یک تحلیلگر مالی در هنگام بررسی میزان بدهی های عمومی چند منطقه شهرداری به نحو زیر گردآوری شده است:

منطقه شهرداری	جمعیت (به هزار نفر)	بدهی سرانه (به ده هزار تومان)	نسبت بدهی به ارزش سرمایه (درصد)
1	457	4.8	6.9
2	1322	5.9	7.7
3	181	3.4	8.3
4	665	5.5	8

- الف) جامعه مورد نظر در این بررسی چیست؟
 ب) در مجموعه داده ها مشخص کنید که کدام متغیر کمی و کدامیک کیفی است؟
 پ) عناصر این مجموعه داده ها، چه چیزی را تشکیل می دهد؟

23

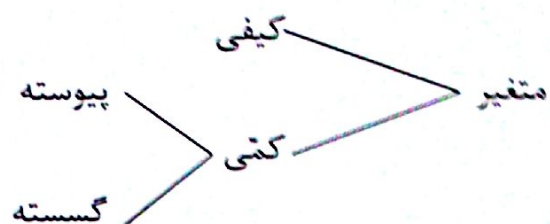
فصل سوم

سطوح سنجش متغیرها

(LEVELS OF MEASUREMENT)

24

• تقسیم بندی مقدماتی پیشین (انواع متغیرها بر حسب ماهیت)



28