

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

درس دوم: سازماندهی دادده‌ها



- آشنایی مقدماتی با شیوه‌های تنظیم و تلخیص داده‌ها و تشکیل جدول فراوانی

هدف‌های رقابتی:



- نحوه مرتب کردن داده‌ها (کمی و رسته‌ای)
- تنظیم جداول فروانی و معیارهای مربوطه
- نمودارهای آماری و کاربرد آنها در توصیف داده‌ها



توصيف اطلاعات

توصیف اطلاعات



(1) انواع طبقه بندی

(2) طبقه بندی داده های کیفی و گسسته

(3) فراوانی ها

(4) طبقه بندی داده های پیوسته

(5) نمودارهای آماری

➤ نمودار دایره ای

➤ نمودار ستونی

➤ نمودار مستطیلی

➤ نمودار چند ضلعی فراوانی

توصیف اطلاعات



- اندازه‌های ثبت شده در مجموعه داده‌ها، اجزای اساسی اطلاعاتی هستند که باید به دقت آنها را مورد بررسی قرار داده و به خوبی توصیف کنیم. به همین دلیل در آمار توصیفی اهداف زیر را دنبال می‌کنیم.

مراحل اساسی توصیف داده‌ها

- خلاصه کردن، دسته‌بندی کردن و تهیه جدول‌های توزیع فراوانی

- بدست آوردن شاخص‌هایی که گرایش به مرکز داده‌ها را معین می‌کنند (شاخصهای مرکزی)

- بدست آوردن شاخص‌هایی که میزان انحراف از مرکز داده‌ها را معلوم می‌کنند (شاخص‌های پراکنگی)

- رسم نمودارهای مناسب برای توصیف بصری داده‌ها

- بررسی کلی نمودارها و پارامترها از لحاظ خصوصیات گرایش به مرکز یا انحراف از مرکز

- اندازه‌های ثبت شده در مجموعه داده‌ها، اجزای اساسی اطلاعاتی هستند که در دسترس محقق قرار دارد. بنابراین در مواجهه با تعداد زیادی داده، ذهن انسان نمی‌تواند محتوای کلی اطلاعات ثبت شده در مجموعه داده‌ها را بلافاصله درک کند. یکی از راه‌های توصیف اطلاعات دسته بندی کردن داده ها است.

جغرافیایی

- طبقه بندی داده ها بر حسب مناطق جغرافیایی مانند کشورها، استان هاو شهرها

زمانی

- طبقه بندی داده ها بر اساس زمان گرد آوری آنها مثل میزان بارندگی ماههای سال(داده های سری زمانی)

کیفی

- طبقه بندی داده ها بر اساس یک متغیر کیفی مانند جنسیت، نژاد و شغل

کمی

- طبقه بندی داده ها بر اساس یک متغیر کمی مانند سن و وزن



در نگاه اول ممکن است داده‌ها اطلاعات مفیدی به ما ندهند و برای اینکه بدانیم این اعداد شامل چه اطلاعاتی هستند، باید آنها را مرتب و دسته‌بندی کنیم تا قابل تفسیر و بهره‌برداری شوند. طبقه‌بندی کردن داده‌های گسسته و کیفی به دلیل محدود بودن پاسخ‌ها بسیار ساده است. در صورتی که طبقه‌بندی متغیرهای پیوسته همواره با پیچیدگی‌های مختلفی روبرو است که در ادامه به آنها نیز خواهیم پرداخت.

چگونه دسته‌بندی کنیم

برای دسته‌بندی داده‌های **کمی** **گسسته** باید هریک از مقادیر متغیر را به همراه فراوانی آنها در یک جدول قرار دهیم.

برای دسته‌بندی داده‌های **کیفی** کافی است فراوانی هر یک از سطوح متغیر کیفی را در یک جدول قرار دهیم

مثال

به چنین جدول‌هایی، جدول، توزیع فراوانی می‌گوییم و میتوان در آنها فراوانی‌های دیگری را نیز تعیین کرد.

طريقه بندي متغير كيفي

طبقه بندي
(کسته)
متغير کمي

فراوانی مطلق

به تعداد داده‌ها در هر طبقه فراوانی مطلق آن طبقه می‌گویند و آن را با f_i نشان می‌دهند.

فراوانی نسبی

اگر فراوانی‌های مطلق را بر کل فراوانی‌ها تقسیم کنیم، فراوانی نسبی (f_i') به دست می‌آید.

$$\frac{f_i'}{n}$$

فراوانی تجمعی

به مجموع فراوانی‌های مطلق طبقه‌های قبل و همان طبقه، فراوانی تجمعی آن طبقه می‌گویند و آن را با cf_i نمایش می‌دهند.

$$cf_n = \sum_{i=1}^n f_i = n$$

فراوانی تجمعی نسبی

می‌توان از تقسیم فراوانی‌های تجمعی بر تعداد داده‌ها، این فراوانی را به دست آورد.

$$cf_i'' = \frac{cf_i}{n}$$

نماینده طبقه

نقطه وسط هر طبقه را نماینده آن طبقه می‌گوییم و آن را با x_i' نشان می‌دهیم.

$$x_i' = \frac{L_i + H_i}{2}$$

توصیف اطلاعات

جدول توزیع فراوانی

$$2+13+18$$

$$\frac{13}{50} = 0.26$$

$$\frac{2}{50} = 0.04$$

$$\frac{43}{50} = 0.86$$

x_i	f_i	cf_i	fi'	cf_i''
0	2	2	0/04	0/04
1	13	15	0/26	0/3
2	18	33	0/36	0/66
3	10	43	0/2	0/86
4	7	50	0/14	1

مثال

جدول توزیع
فراوانی همراه با
فراوانیهای مطلق،
تجمعی، نسبی،
تجمعی نسبی برای
تعداد فرزندان ۵۰
کارمند در مثال قبل

طبقه بندی داده های پیوسته

اگر داده ها از یک صفت کمی پیوسته اندازه گیری شده باشند، لزوماً تکرار نخواهند داشت بنابراین باید آنها را به روش زیر طبقه بندی کرد.

۱- **تعدادی طبقه (K)** را به تناسب تعداد داده ها (n) و یا با استفاده از رابطه زیر تعیین کنید.

$$k = 1 + \sqrt[3]{\log(n)}$$

(دستور استورجس)

۲- **دامنه تغییرات** داده ها (R) را از رابطه زیر محاسبه کنید.

بزرگترین داده

$$R = \text{Max}(x_i) - \text{Min}(x_i)$$

کوچکترین داده

۳- **فاصله طبقه ها (C)** را از رابطه زیر به دست آورید:
(با تقریب اضافی)

$$c = R/k$$

توصیف اطلاعات

طبقه بندی داده ها



۴- حدود هر طبقه (Li - Hi) را تعیین کنید و داده‌هایی که در هر طبقه قرار می‌گیرند، شمارش کرده و به عنوان فراوانی (fi) آن طبقه ثبت کنید.

مثال

معدل ۵۰ دانشجو با تقریب تا یک رقم اعشار، به شرح زیر است. این داده ها را به ۸ طبقه تقسیم بندی کنید و فراوانی های مطلق نسبی و تجمعی و تجمعی نسبی را به دست آورید.

برای طبقه بندی این داده‌ها،
مراحل دسته بندی کردن
داده‌های کمی پیوسته را پی
می‌گیریم

[2/9]	1/5	1/8	2/4	2/2	2/1	2/2	1/6	1/9	2/1
2/6	2/1	2/5	2/0	2/3	2/3	1/7	1/8	2/3	1/8
1/4	2/6	2/2	1/9	2/0	1/7	1/7	1/9	[1/4]	1/8
2/0	2/0	2/0	2/5	2/2	2/2	1/9	1/8	2/4	2/9
1/4	1/6	1/9	2/9	2/4	1/6	1/8	1/9	2/5	2/1

[بزرگترین و کوچکترین عدد در بین داده ها را مشخص کرده ایم]

توصیف اطلاعات

طبقه بندی داده ها

پهار مرحله طبقه بندی داده های پیوسته برای این مثال

حدود طبقات	x'_i	f_i	fi'	cf_i	$cf_{i''}$
1/4 - 1/6	1/5	4	0/08	4	8/0
1/6 - 1/8	1/7	6	0/12	10	20/0
1/8 - 2	1/9	12	0/24	22	44/0
2 - 2/2	2/1	9	0/18	31	62/0
2/2 - 2/4	2/3	8	0/16	39	78/0
2/4 - 2/6	2/5	6	0/12	45	90/0
2/6 - 2/8	2/7	2	0/04	47	94/0
2/8 - 3	2/9	3	0/06	50	100
جمع		50	1/00	—	—

۱- تعداد طبقه در صورت

مسئله، ۸ تعیین شده است

۲- دامنه تغییرات را

$$2/9 - 1/4 = 1/5$$

به دست می آوریم

۳- فاصله طبقات از رابطه زیر

(C) و با تقریب اضافه ۰/۲ به

دست آمده است.

۴- حدود طبقات را با شروع از

کمترین مقدار تعیین کرده ایم.

$$c = \frac{2.9 - 1.4}{8} = 0.2$$

معمولاً برای توصیف بهتر داده‌های آماری از نمودارها استفاده می‌کنیم. عموماً این نمودارها در ارتباط با داده‌های پیوسته به کار گرفته می‌شوند.

منظور از نمایش نموداری داده‌ها، تجسم عینی اطلاعات نهفته در آنها است.

تعدادی از نمودارهای مهم آماری

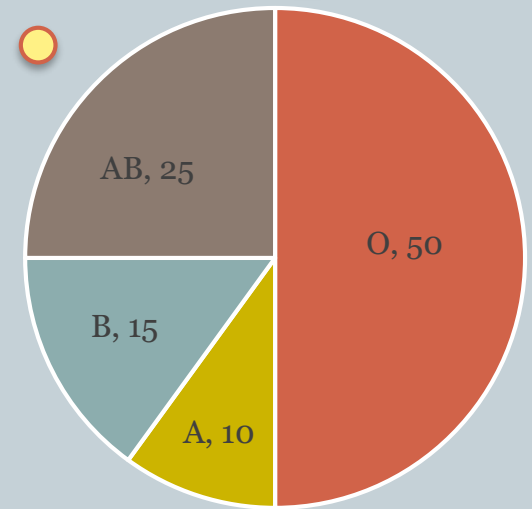
- ۱- نمودار دایره‌ای (معمولاً برای داده‌های کیفی استفاده می‌شود)
- ۲- نمودار ستونی (معمولاً برای داده‌های کمی گسسته استفاده می‌شود)
- ۳- نمودار مستطیلی یا هیستوگرام (کمی پیوسته)
- ۴- نمودار چند ضلعی فراوانی (کمی پیوسته)
- ۵- نمودار ساقه و برگ (کمی گسسته و پیوسته)
- ۶- نمودار جعبه‌ای (کمی پیوسته)

نحوه رسم نمودار دایره‌ای را در این مثال ببینید.

مثال گروه خون ۱۰۰ بیمار بستری در یک بیمارستان را در اختیار داریم نمودار دایره‌ای داده‌ها را رسم کنید.

$$0/15 \times 360 = 54$$

گروه خون	فراوانی f_i	فراوانی نسبی f_i'	قطاع دایره (درجه)
O	50	0/5	180
A	10	0/1	36
B	15	0/15	54
AB	25	0/25	90

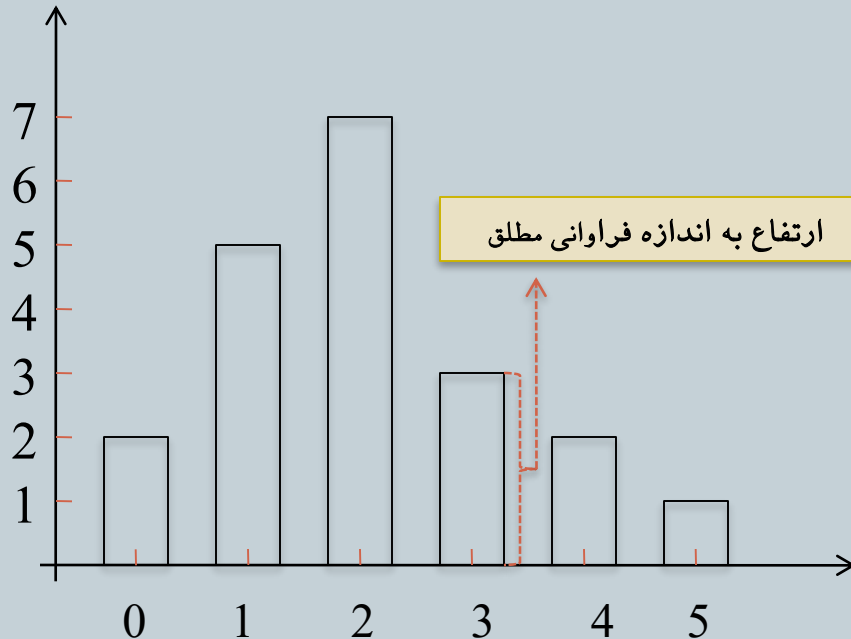


روش رسم نمودار ستونی

مثال

برای داده های زیر نمودار ستونی رسم کنید.

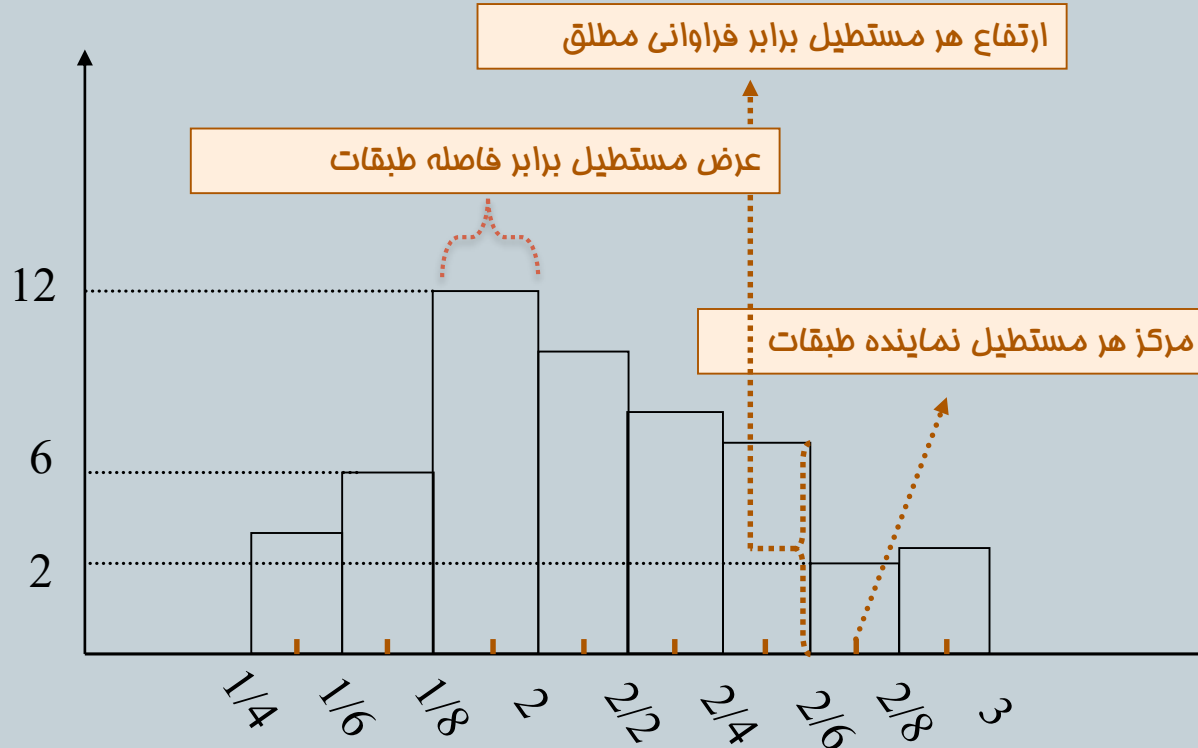
3 1 2 2 0 1 2 3 2 1 1 5 2 0 1 2 2 4 4 3



برای رسم نمودار ستونی کافی است مقادیر متغیر را روی محور افقی تعیین کنید سپس مستطیل هایی با فاصله و به ارتفاع فراوانی مطلق برای هر مقدار متغیر رسم کنید.

ابتدا جدول توزیع فراوانی را تشکیل می دهیم

x	f_i
0	2
1	5
2	7
3	3
4	2
5	1



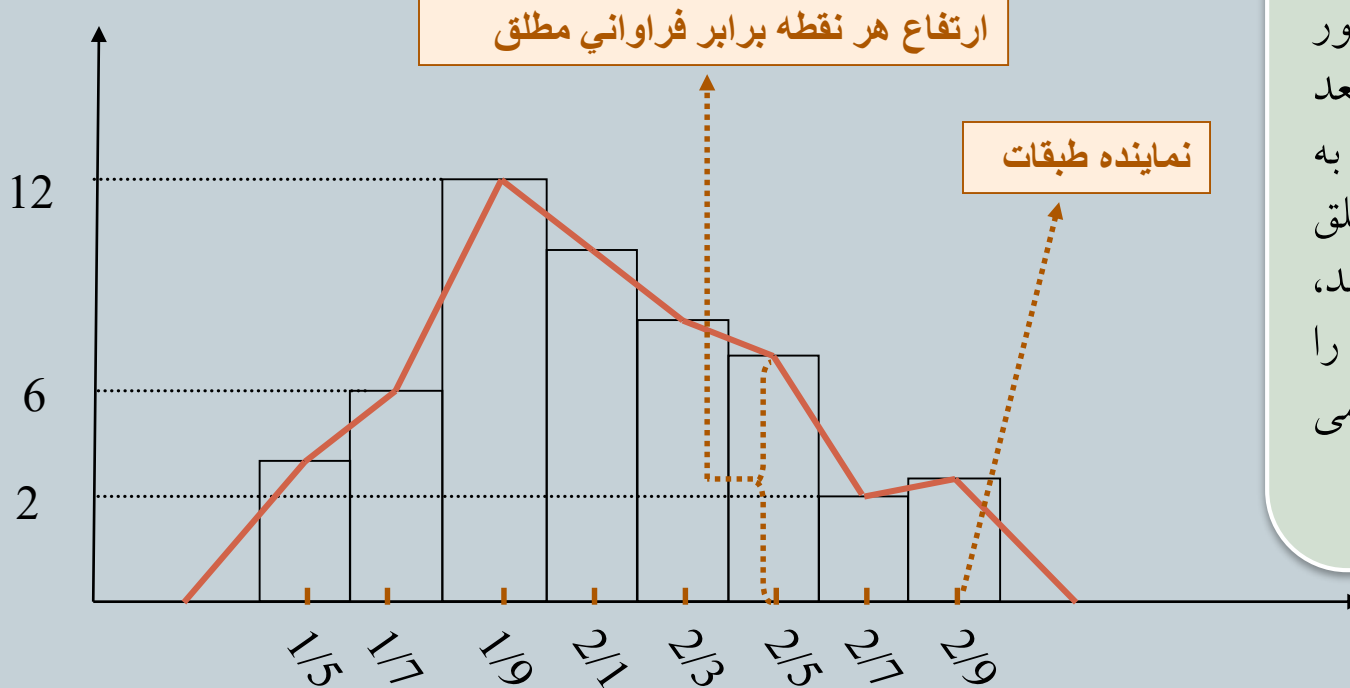
برای رسم هیستوگرام فراوانی یا نمودار مستطیلی داده ها ابتدا حدود طبقات را روی محور افقی تعیین می کنیم. بعد مستطیل هایی که عرض آنها به اندازه فاصله طبقات و ارتفاع آنها به اندازه فراوانی مطلق طبقه مربوطه باشد، رسم می کنیم.

نمودار چند ضلعی

توصیف اطلاعات

روش رسم نمودار چند ضلعی

مستطیلها را روی این نمودار
برای درک بهتر نمودار چند
ضلعی قرار داده ایم

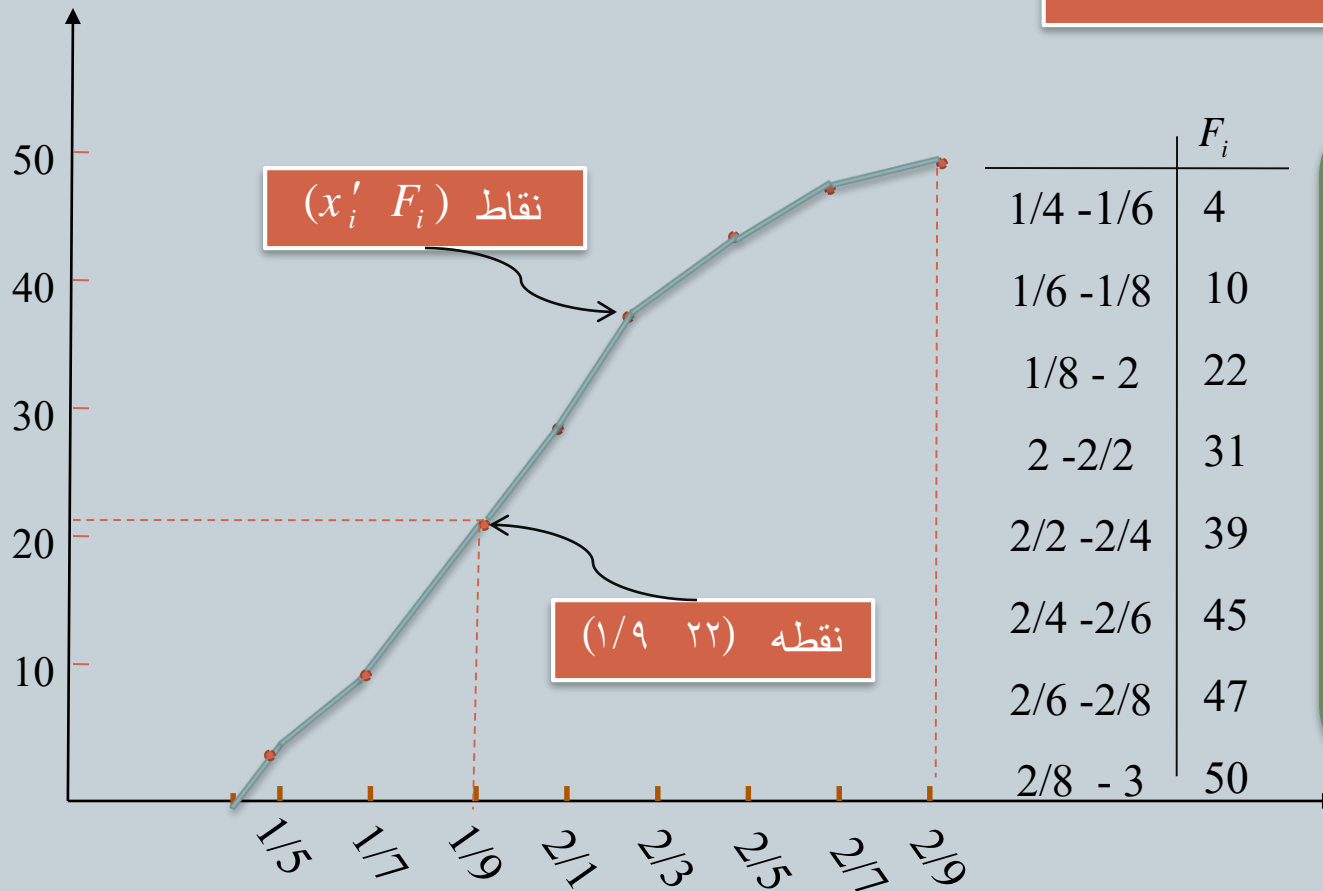


برای رسم چند ضلعی
فراوانی داده ها نماینده
طبقات را روی محور
افقی تعیین می کنیم. بعد
نقاطی که ارتفاع آنها به
اندازه فراوانی مطلق
طبقه مربوطه باشد،
مشخص کرده و آنها را
به یکدیگر متصل می
کنیم.

نمودار چند ضلعی

توصیف اطلاعات

روش رسم نمودار چند ضلعی فراوانی تجمعی



برای رسم چند ضلعی فراوانی تجمعی داده‌ها نماینده طبقات را روی محور افقی تعیین می‌کنیم. بعد نقاطی که ارتفاع آنها به اندازه فراوانی تجمعی طبقه مربوطه باشد، مشخص کرده و آنها را به یکدیگر متصل می‌کنیم.