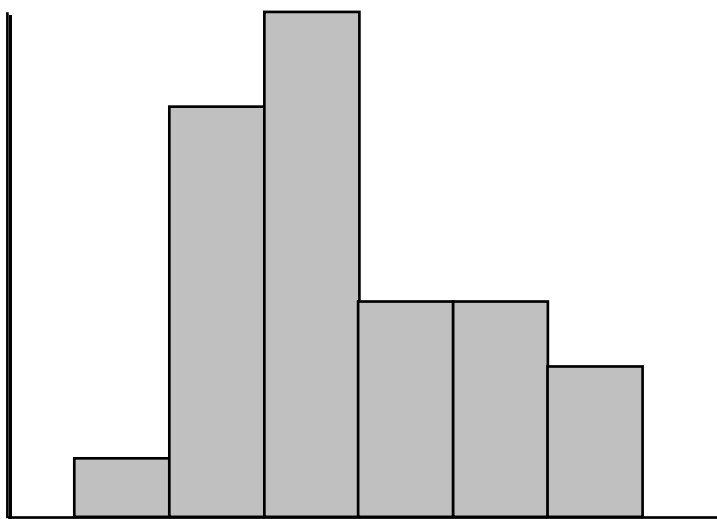


آمار توصیفی

روشهای جدولی و نموداری

آمار توصیفی

روشهای جدولی و نموداری



- خلاصه کردن داده های کیفی
- خلاصه کردن داده های کمی
- تحلیل کاوشگرانه داده ها
- جدولهای توافقی و نمودارهای پراکنش

خلاصه کردن داده های کیفی

- توزیع فراوانی
- توزیع فراوانی نسبی
- توزیع درصد فراوانی
- نمودار میله ای
- نمودار دایره ای

توزیع فراوانی

- توزیع فراوانی خلاصه جدولی از داده هایی است که فراوانی (یا تعداد) موارد هریک از رده های نامتداخل را نشان می دهد.
- هدف، فراهم کردن نگرشی درباره داده هایی است که نمی توان تنها با مشاهده داده های اصلی به آنها دست یافت.

مثال: مسافر خانه

مهمانان مقیم مسافر خانه مایلند کیفیت منازلشان را در پنج سطح عالی، خیلی خوب، متوسط، ضعیف و بد رده بندی کنند. این رده بندی با نمونه گیری از ۲۰ مسافر تهیه شد که در زیر نشان داده شده است.

خیلی خوب	متوسط	ضعیف
خیلی خوب	خیلی خوب	خیلی خوب
ضعیف	ضعیف	خیلی خوب
بد	متوسط	بد
ضعیف	عالی	ضعیف
متوسط	ضعیف	متوسط
متوسط	ضعیف	متوسط

مثال: مسافر خانه

- توزیع فراوانی

<u>درجه بندی</u>	<u>فراوانی</u>
بد	2
ضعیف	3
متوسط	5
خیلی خوب	9
عالی	1
جمع کل	20

توزیع فراوانی نسبی

- فراوانی نسبی هر رده تابع یا احتمال تعداد کل داده های متعلق به رده است.
- توزیع فراوانی نسبی خلاصه جدولی از مجموعه داده هایی است که فراوانی نسبی هر رده را نشان می دهد.

توزیع درصد فراوانی

- درصد فراوانی هر رده حاصل ضرب فراوانی نسبی در 100 است.
- توزیع درصد فراوانی خلاصه جدولی مجموعه داده هایی است که درصد فراوانی نسبی را نشان می دهد.

مثال: مسافر خانه

- توزیع فراوانی نسبی و درصد فراوانی

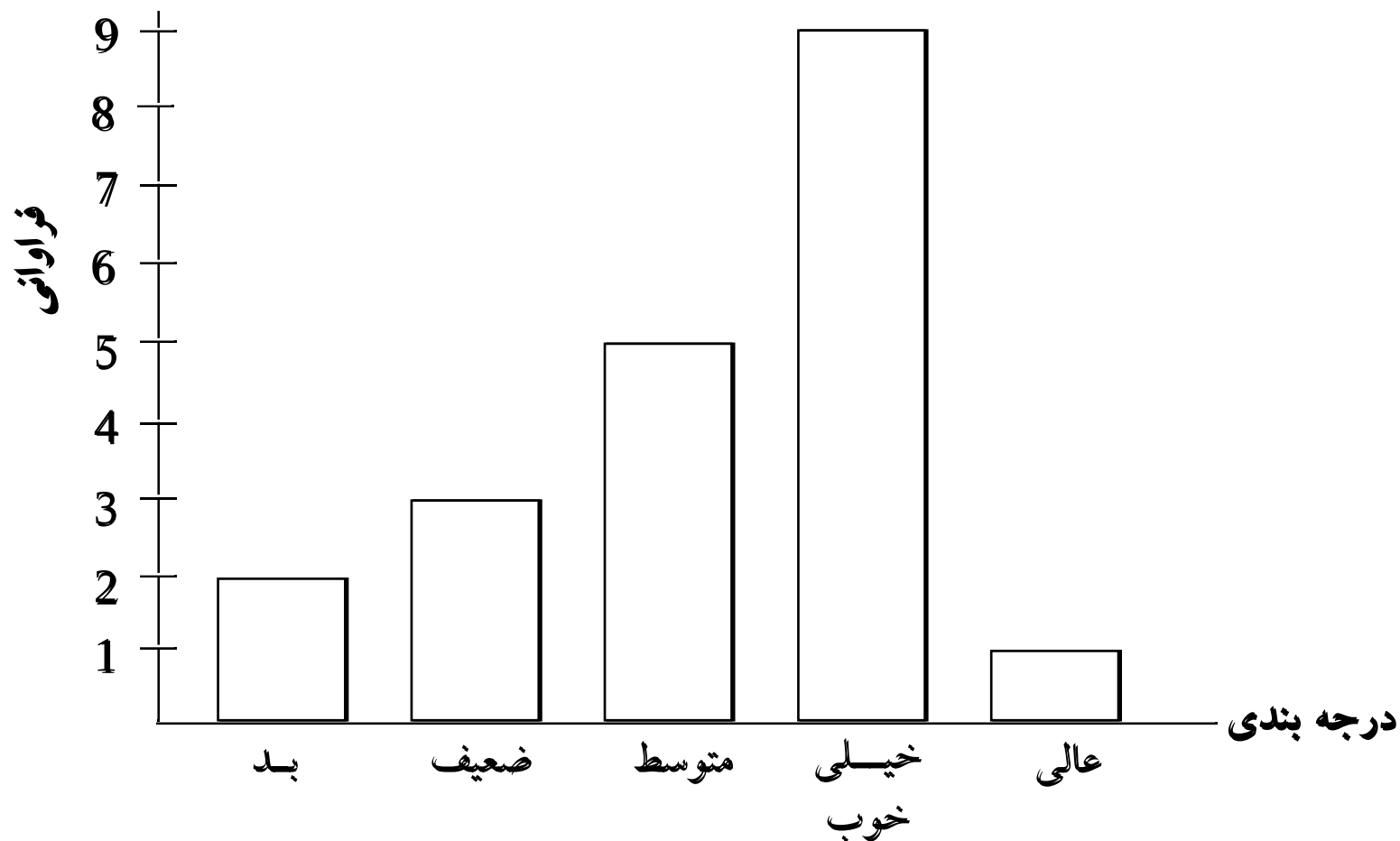
درجه بندی	نسبی	درصد
بد	0.10	10
ضعیف	0.15	15
متوسط	0.25	25
خیلی خوب	0.45	45
عالی	0.05	5
کل	1.00	100

نمودار میله ای

- نمودار میله ای یک وسیله گرافیکی برای نمایش داده های کیفی است.
- در محور افقی برچسب های هر یک از رده ها را مشخص می کنیم.
- یکی از مقیاس های فراوانی، فراوانی نسبی یا درصد فراوانی در محور عمودی به کار می رود.
- با رسم میله ای با پهنای ثابت در بالای هر برچسب رده، ارتفاع مناسب را نشان می دهیم.
- برای اینکه مشخص شود که هر رده یک رسته جداست، میله ها را از هم جدا می کنیم.

مثال: مسافر خانه

• نمودار دایره ای

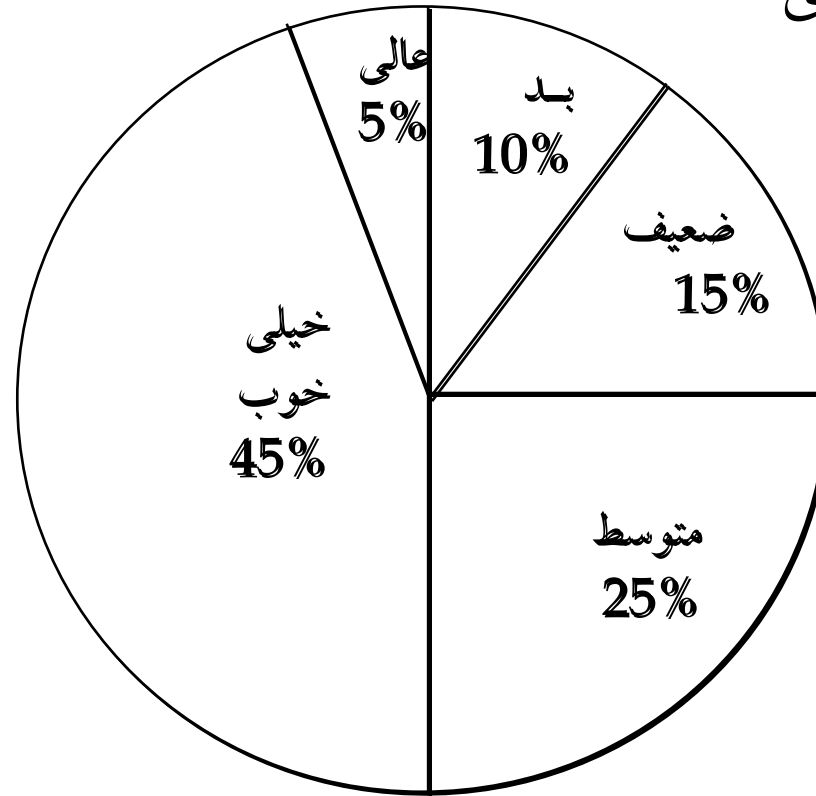


نمودار دایره ای

- نمودار دایره ای یک وسیله گرافیکی معمول برای معرفی توزیع های فراوانی نسبی داده های کیفی است.
- ابتدا دایره ای رسم کنید سپس آن را به قطاع هایی متناظر با فراوانی نسبی هر رده تقسیم کنید.
- چون دایره 360 درجه است، یک رده با فراوانی نسبی 0.25:
 $90 = (360)0.25$ درجه دایره را به خود اختصاص می دهد.

مثال: مسافر خانه

• نمودار دایره ای



رتبه بندی کیفیت

مثال: مسافرخانه

- نتایج حاصل از نمودار دایره ای

– نصف مشتری ها به کیفیت مسافرخانه رتبه خیلی خوب و عالی داده اند (به سمت چپ دایره دقت کنید) این مساله برای مدیر خوشایند است.

– تنها یک مشتری رتبه عالی و دو مشتری رتبه بد داده است (به بالای دایره دقت کنید) این مساله باید برای مدیر ناخوشایند باشد.

خلاصه داده های کمی

- توزیع فراوانی
- توزیع های فراوانی نسبی و درصد فراوانی
- نمودار نقطه ای
- هیستوگرام
- توزیع های تجمعی
- تاکدیس (Ogive)

مثال: تعمیرگاه اتومبیل

مدیر تعمیرگاه اتومبیل مایل به تهیه تصویر خوبی از توزیع هزینه های قطعات تنظیم موتور است. یک نمونه 50 تایی از صورت حساب مصرف کنندگان در لیست زیر آمده است که در آن ارقام به نزدیکترین عدد گرد شده است:

91	78	93	57	75	52	99	80	97	62
71	69	72	89	66	75	79	75	72	76
104	74	62	68	97	105	77	65	80	109
85	97	88	68	83	68	71	69	67	74
62	82	98	101	79	105	79	69	62	73

توزیع فراوانی

- پیشنهاداتی برای انتخاب تعداد رده ها
 - بین 5 و 20 رده انتخاب کنید.
 - مجموعه داده های با تعداد عناصر زیاد معمولا تعداد رده های بیشتری لازم دارند.
 - مجموعه داده های با تعداد عناصر کم معمولا تعداد رده های کمتری لازم دارند.

توزیع فراوانی

- پیشنهاداتی برای انتخاب عرض رده ها
 - رده هایی با پهنای مساوی را به کار ببرید.
 - پهنای تقریبی رده

بیشترین مقدار داده – کمترین مقدار داده
تعداد رده ها

مثال: تعمیرگاه اتومبیل

- توزیع فراوانی

اگر عرده را در نظر بگیریم:

$$\text{پهنای تقریبی رده} = (109 - 52)/6 = 9.5 \cong 10$$

<u>هزینه</u>	<u>فراوانی</u>
50-59	2
60-69	13
70-79	16
80-89	7
90-99	7
100-109	<u>5</u>
جمع	50

مثال: تعمیرگاه اتومبیل

- توزیع فراوانی نسبی و درصد فراوانی

هزینه	فراوانی <u>نسبی</u>	درصد <u>فراوانی</u>
50-59	0.04	4
60-69	0.26	26
70-79	0.32	32
80-89	0.14	14
90-99	0.14	14
100-109	<u>0.10</u>	<u>10</u>
جمع	1.00	100

مثال: تعمیرگاه اتومبیل

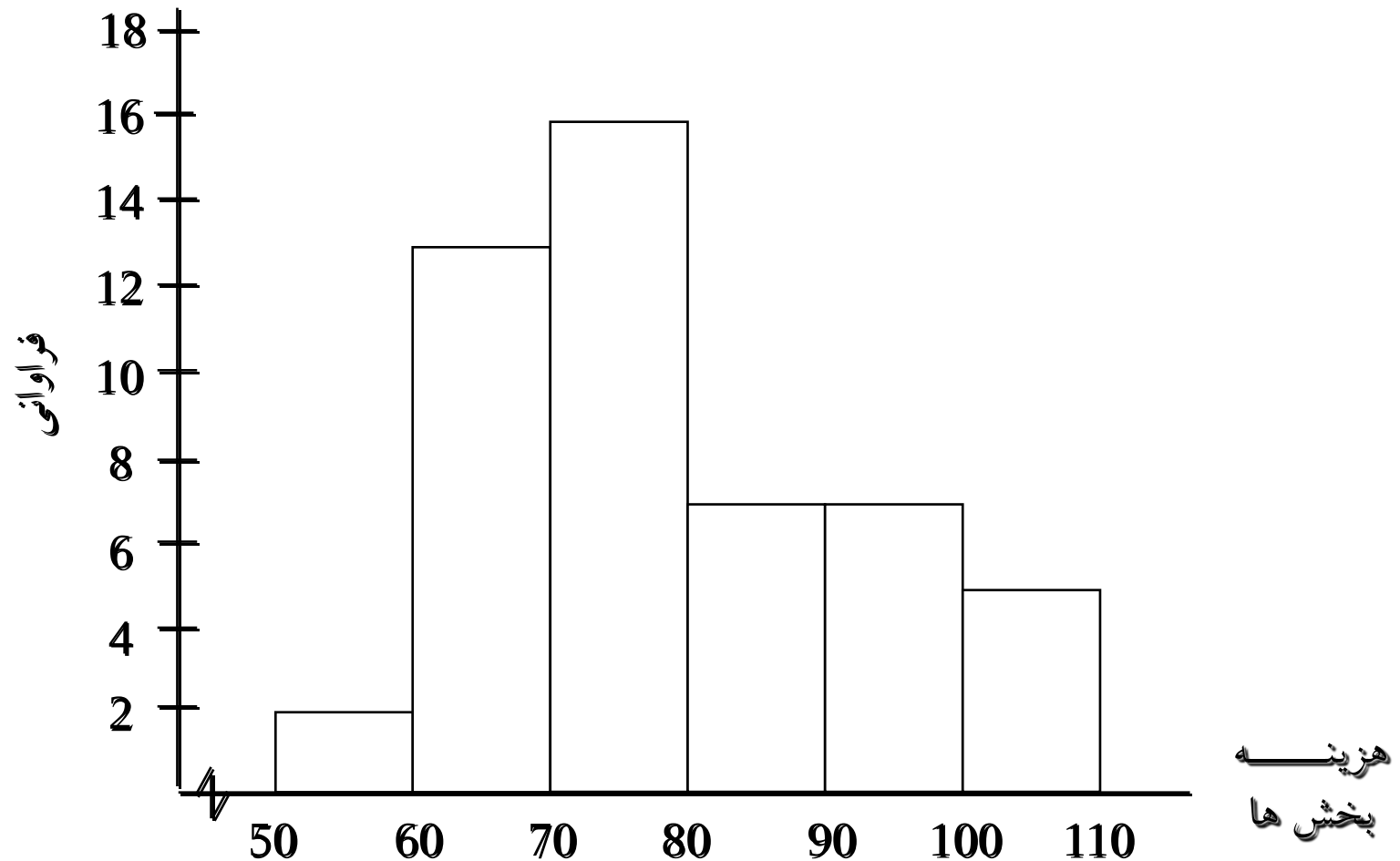
- نتایج حاصل از توزیع درصد فراوانی
 - تنها 4% از هزینه قطعات در رده 50 - 59 واقع است.
 - 30% از هزینه قطعات کمتر از 70 است.
 - بیشترین درصد (32% یا نزدیک یک سوم) از هزینه قطعات در رده 70 - 79 است.
 - 10% از هزینه قطعات 100 یا بیشتر از 100 است.

هیستوگرام

- نمایش گرافیکی معمول دیگر برای داده های کمی، هیستوگرام است.
- متغیر مورد نظر در محور افقی قرار داده شده است.
- بالای هر رده مستطیلی با ارتفاع متناسب با فراوانی، فراوانی نسبی، یا درصد فراوانی فاصله رسم می شود.
- برخلاف نمودار میله ای، در هیستوگرام، مستطیلهای رده های مجاور از هم جدا نیستند.

مثال: تعمیرگاه اتومبیل

• هیستوگرام



توزیع های تجمعی

- توزیعهای فراوانی تجمعی: تعداد داده های کمتر یا مساوی حد بالای هر رده را نشان می دهد.
- توزیعهای فراوانی نسبی تجمعی: نسبت داده های کمتر یا مساوی حد بالای هر رده را نشان می دهد.
- توزیعهای درصد فراوانی تجمعی: درصد داده های کمتر یا مساوی حد بالای هر رده را نشان می دهد.

مثال: تعمیرگاه اتومبیل

- توزیع های تجمعی

قیمت	فراوانی تجمعی	فراوانی نسبی تجمعی	درصد فراوانی تجمعی
≤ 59	2	0.04	4
≤ 69	15	0.30	30
≤ 79	31	0.62	62
≤ 89	38	0.76	76
≤ 99	45	0.90	90
≤ 109	50	1.00	100

تاقدیس

- تاقدیس یک نمودار توزیع تجمعی است.
- مقادیر داده ها در محور افقی نشان داده می شوند.
- در محور عمودی:
 - فراوانی های تجمعی، یا
 - فراوانی های نسبی تجمعی، یا
 - درصد فراوانی های تجمعی نمایش داده می شود.
- (هریک از) فراوانی های هر رده به شکل نقطه رسم می شوند.
- نقاط رسم شده با خطوط مستقیم به یکدیگر متصل می شوند.

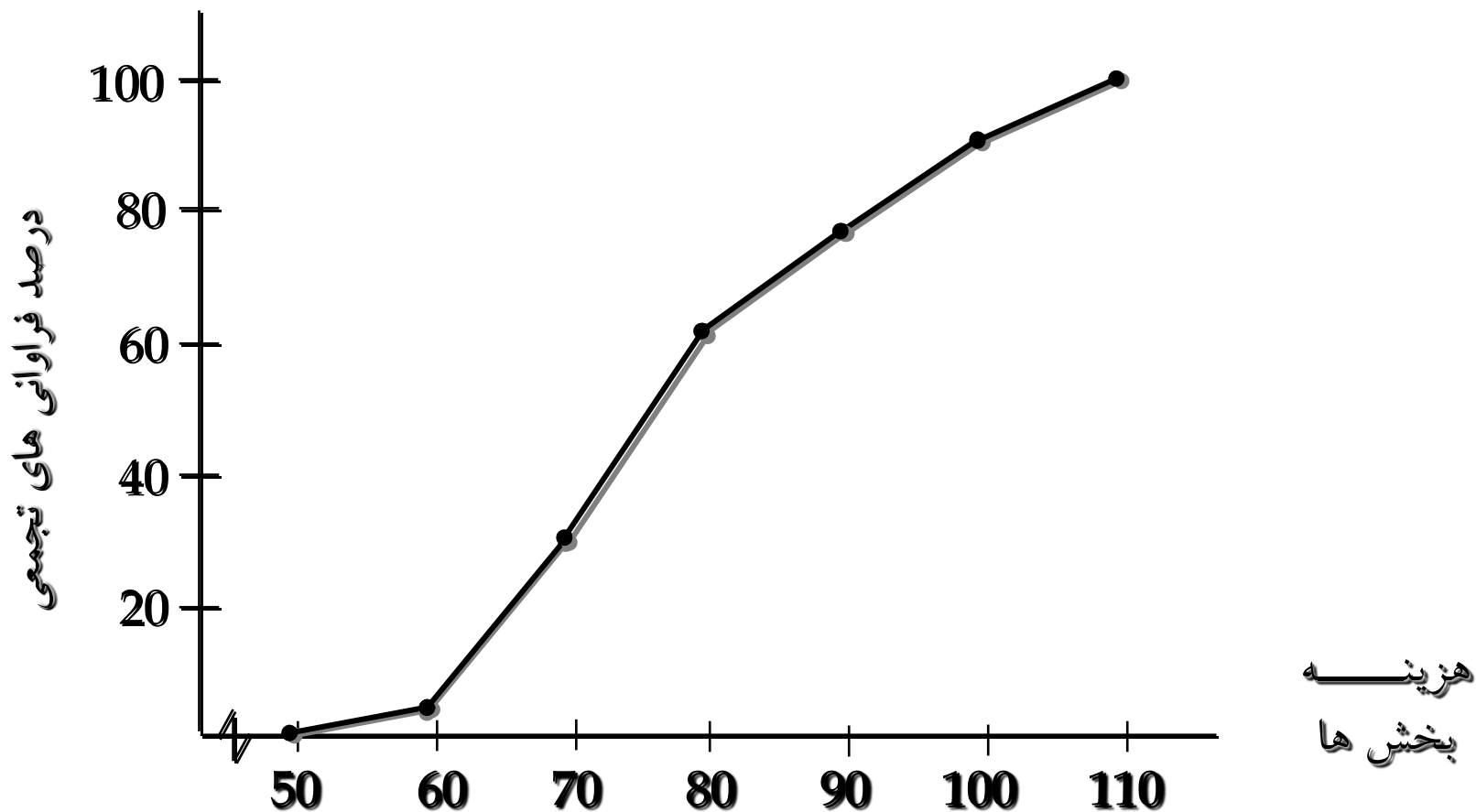
مثال: تعمیرگاه اتومبیل

- تاقدیس

- چون حدود رده داده های هزینه بخشها به صورت 59 - 50 و 69 - 60 و غیره است، یک شکاف بین 59 و 60، 69 و 70 و غیره ایجاد می شود.
- این شکاف ها با نقاط رسم شده در وسط فاصله حدود رده ها برطرف می شوند.
- بنابراین 59.5 در رده 59 - 50 و 69.5 در رده 69 - 60 و بقیه به همین ترتیب استفاده می شود.

مثال: تعمیرگاه اتومبیل

- تاقدیس درصد فراوانی های تجمعی



تحلیل کاوشگرانه داده ها

- روشهای تحلیل کاوشگرانه داده ها شامل عملیات حسابی ساده و تصاویر ساده برای رسم است که میتواند در خلاصه کردن سریع داده ها به کار رود.
- یکی از آن روش ها نمایش ساقه و برگ است.

نمایش ساقه و برگ

- نمایش ساقه و برگ هم ترتیب رتبه ای و هم شکل توزیع داده ها را نشان میدهد.
- نمایش ساقه و برگ مشابه هیستوگرام است، ولی به جهت نشان دادن مقادیر واقعی داده ها از آن برتر است.
- اولین رقم هر کدام از داده ها در سمت چپ یک خط عمودی مرتب می شوند.
- در سمت راست خط، آخرین رقم هر مورد در ترتیب رتبه ای را ثبت میکنیم.
- هر خط در نمایش، یک ساقه است.
- هر رقم در ساقه، یک برگ است.

مثال: تعمیر گاہ اتومبیل

- نمایش ساقه و برگ

[illegible]

نمایش ساقه - برگ مبسوط

- اگر گمان میکنیم که نمایش ساقه - برگ اصلی، داده ها را خیلی خلاصه کرده است، می توانیم نمایش را با دو ساقه اضافی برای هر رقم مقدم بسط دهیم.
- به هر حال یک مقدار ساقه دو چیز را شرح می دهد، اولین مقدار مساوی مقادیر بازه ۰-۴ و دومین مقدار مساوی مقادیر ۵-۹ است.

مثال: تعمیرگاه اتومبیل

- نمایش ساقه - برگ مبسوط

5	2								
5	7								
6	2	2	2	2					
6	5	6	7	8	8	8	9	9	9
7	1	1	2	2	3	4	4		
7	5	5	5	6	7	8	9	9	9
8	0	0	2	3					
8	5	8	9						
9	1	3							
9	7	7	7	8	9				
10	1	4							
10	5	5	9						

نمایش ساقه و برگ

- واحدهای برگ

- برای نمایش هر برگ از یک رقم استفاده می شود.
- در مثال قبل، واحد برگ 1 بود.
- واحدهای برگ می تواند 100 ، 10 ، 1 ، 0.1 و غیره باشد.
- درجایی که واحد برگ نشان داده نشده، آن را مساوی 1 می گیریم.

مثال: واحد برگ = 0.1

اگر داده هایی با مقادیر زیر را داده باشیم

8.8 11.0 10.2 9.1 9.4 11.7 8.6

نمودار ساقه و برگ این داده ها به شکل زیر است:

واحد برگ = 0.1

8		6	8
9		1	4
10			2
11		0	7

مثال: واحد برگ = 10

اگر داده هایی با مقادیر زیر را داده باشیم

1806 1717 1974 1791 1682 1910 1838

نمودار ساقه و برگ این داده ها به شکل زیر است:

واحد برگ = 10

16	8	
17	1	9
18	0	3
19	1	7

جدول های توافقی و نمودار پراکنش

- تا اینجا روی روشهایی که برای خلاصه کردن داده های یک متغیره در یک زمان استفاده می شود، تمرکز کردیم.
- اغلب یک مدیر به روشهای جدولی و گرافیکی که در فهم رابطه بین دو متغیر کمک می کنند، علاقمند است.
- جدول توافقی و نمودار پراکنش دو روش برای خلاصه کردن همزمان داده های دو (یا چند) متغیره است.

جدول توافقی

- جدول توافقی یک روش جدولی از خلاصه داده های دو متغیر در یک زمان است.
- جدول توافقی موقعی قابل استفاده است که:
 - یک متغیر کیفی و دیگری کمی باشد
 - هر دو متغیر کیفی باشد
 - هر دو متغیر کمی باشد
- برچسب های سمت چپ و بالای حاشیه، رده های دو متغیر را معین می کند.

مثال: خانه های ساحلی

- جدولبندی متقاطع

تعداد خانه های ساحلی فروخته شده طی دو سال گذشته از هر مدل

و قیمت در زیر نشان داده شده است:

دامنه قیمت	مدل خانه				جمع
	مستعمرات	مزرعه	دوتکه	ساختمان	
$\leq 99,000$	18	6	19	12	55
$> 99,000$	12	14	16	3	45
جمع	30	20	35	15	100

مثال: خانه های ساحلی

- نتایج حاصل از جدول توافقی پیشین
 - بیشترین تعداد خانه ها از نوع دو تکه و ارزانتر یا مساوی با 99000 است.
 - تنها سه خانه در نمونه از نوع ساختمان و گرانتر از 99000 است.

جدول توافقی: درصد ستونها و ردیفها

- تبدیل ورودی های جدول به درصد ردیفها و درصد ستونها دید
بهتری درباره رابطه بین دو متغیر به ما می دهد.

مثال: خانه های ساحلی

- درصد ردیفها

دامنه قیمت	مدل خانه				جمع
	مستعمرات	مزرعه	دوتکه	ساختمان	
$\leq 99,000$	32.73	10.91	34.55	21.82	100
$> 99,000$	26.67	31.11	35.56	6.67	100

دقت: در واقع جمع ردیفها 100.01 است که گرد شده اند.

مثال: خانه های ساحلی

• درصد ستونها

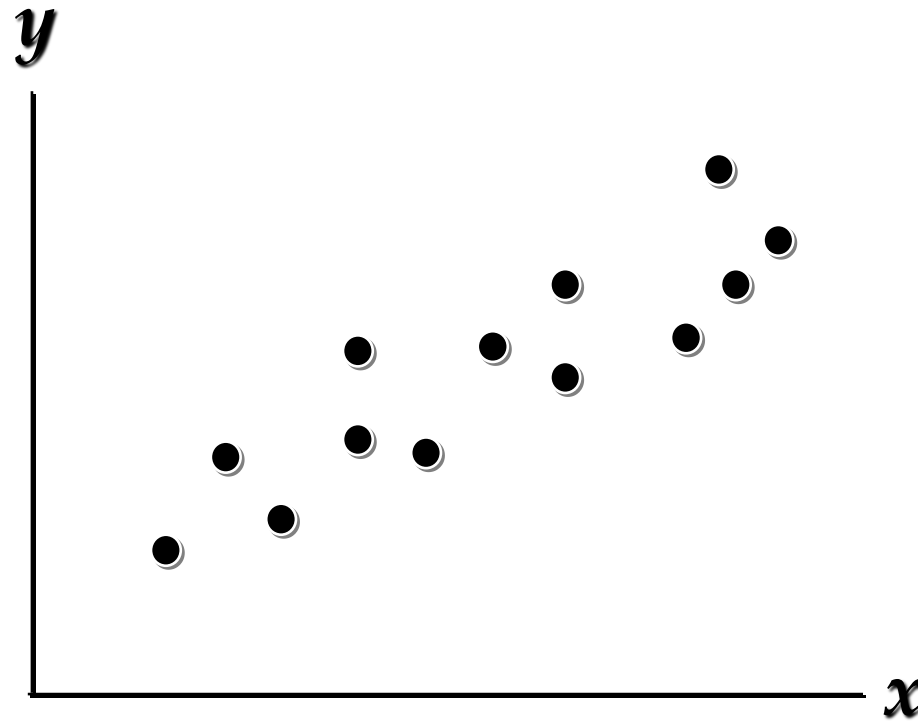
دامنه قیمت	مدل خانه			
	مستعمرات	مزرعه	دوتکه	ساختمان
$\leq 99,000$	60.00	30.00	54.29	80.00
$> 99,000$	40.00	70.00	45.71	20.00
جمع	100	100	100	100

نمودار پراکنش

- نمودار پراکنش نمایش نموداری رابطه بین دو متغیر کمی است.
- یک متغیر در محور افقی متغیر دیگر در محور عمودی نشان داده شده است.
- طرح کلی نقاط رسم شده، هر گونه رابطه بین متغیرها را نشان می دهد.

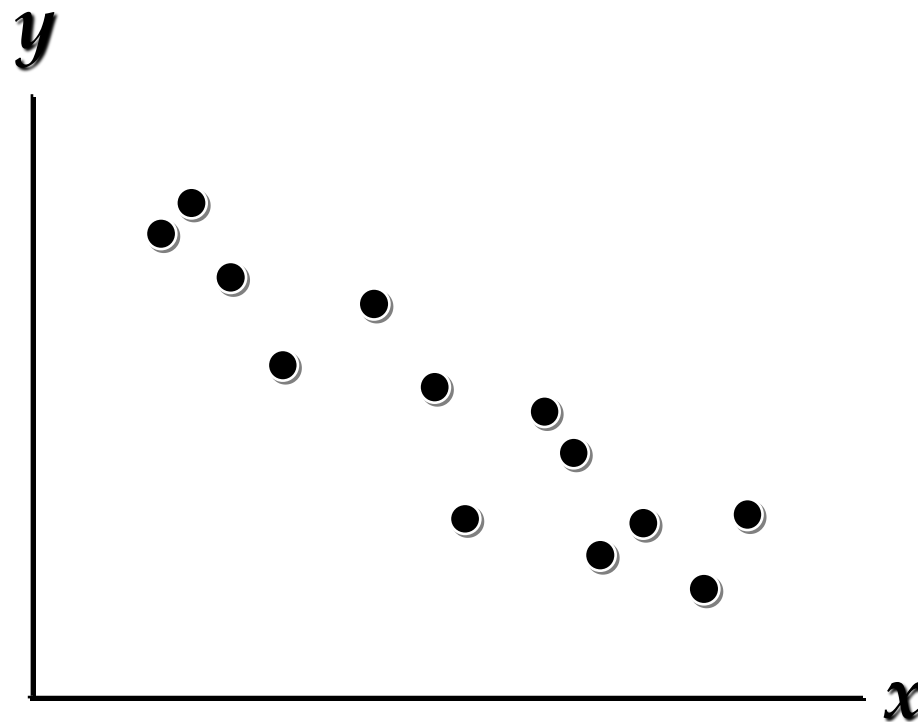
نمودار پراکنش

• رابطه مثبت



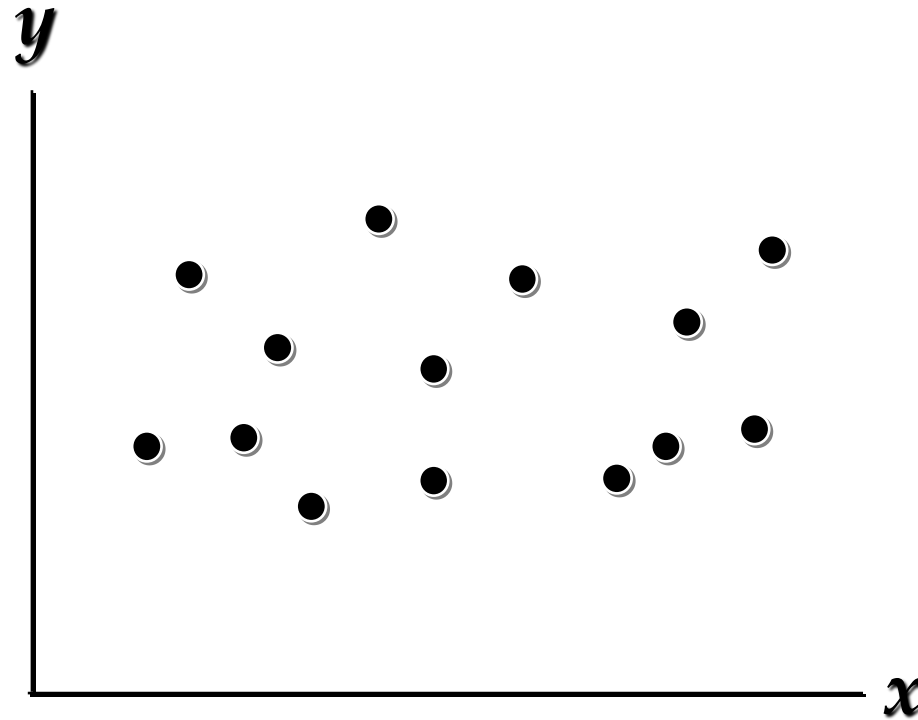
نمودار پراکنش

- رابطه منفی



نمودار پراکنش

- رابطه غیر ظاهری



مثال: تیم فوتبال

- نمودار پراکنش

تیم فوتبالی علاقه مند است درباره وجود هر گونه رابطه بین دفاع و امتیاز کسب شده تحقیق کند.

$x =$ تعداد دفاع

1

3

2

1

3

$y =$ تعداد پیروزی ها

14

24

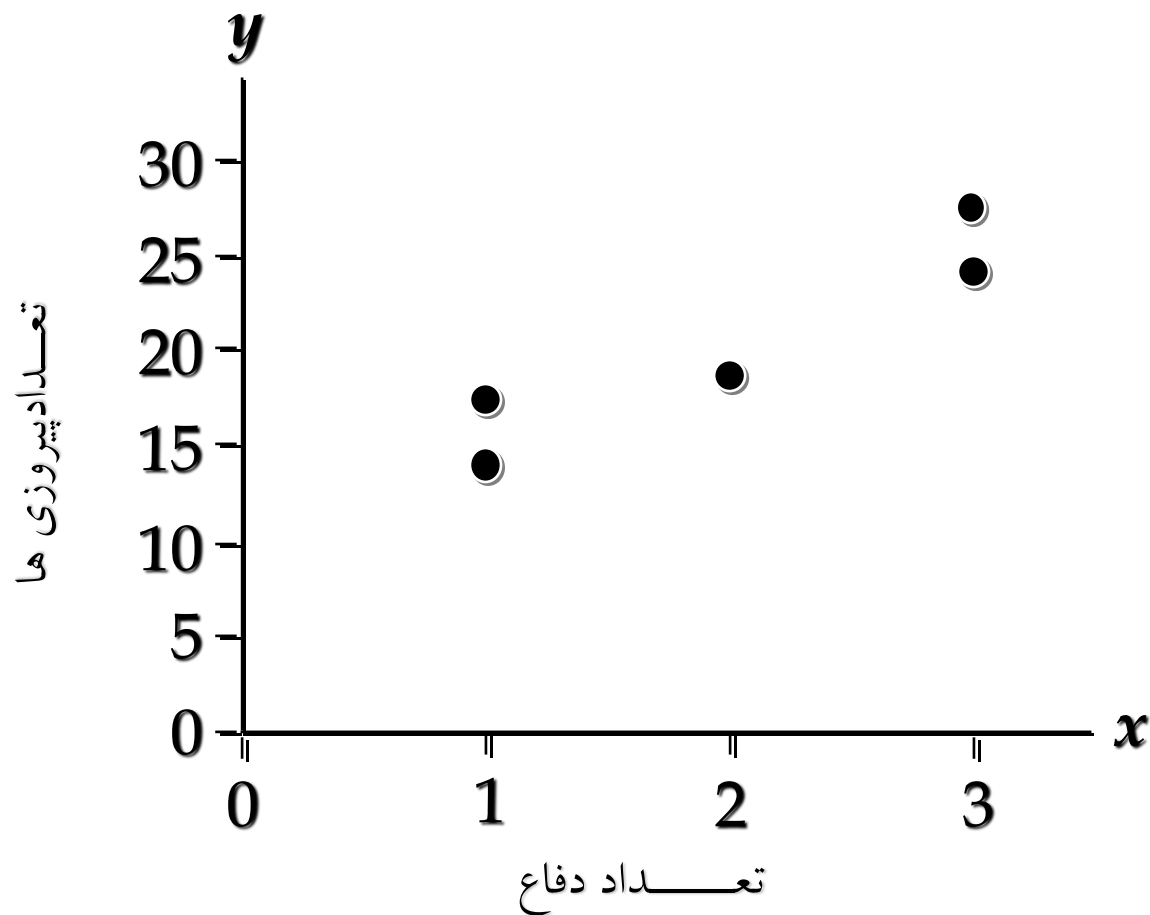
18

17

27

مثال: تیم فوتبال

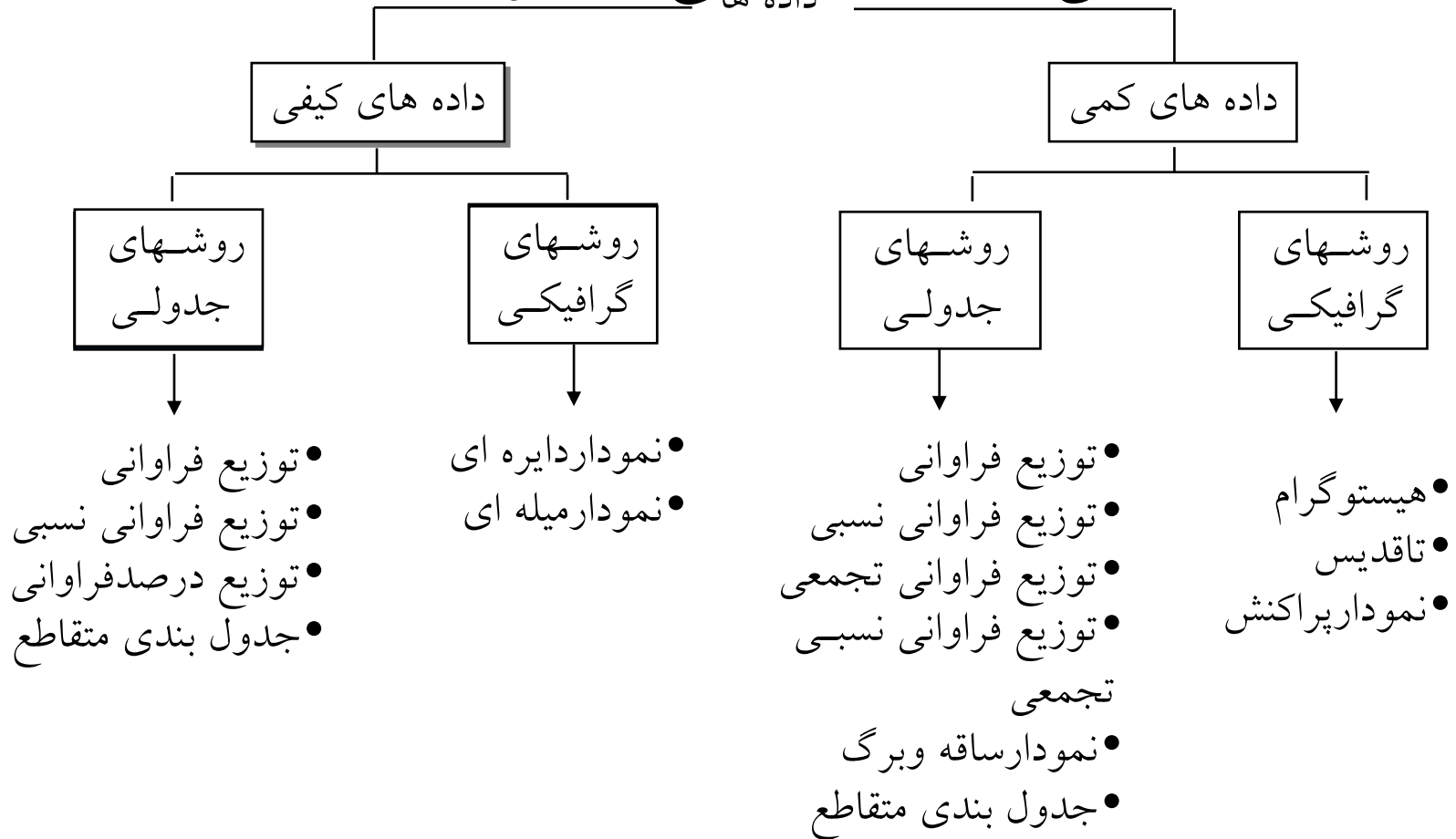
• نمودار پراکنش



مثال: تیم فوتبال

- نمودار پراکنش قبلی یک رابطه مثبت بین تعداد دفاع ها و تعداد نقاط پیروزی ها را نشان می دهد.
- امتیازهای بالاتر وابسته به تعداد دفاع های بالاتر است.
- رابطه کامل نیست، تمام نقاط رسم شده در نمودار پراکنش روی یک خط مستقیم نیستند.

روش های جدولی و نموداری داده ها



پایان

