

شاخص های مرکزی و پراکندگی

شاخص‌های مرکزی:

میانگین:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n f_i x_i$$

مهم‌ترین و مفیدترین شاخص مرکزی
متوسط (معدل) مقادیر داده‌ها
جمع کل داده‌ها تقسیم بر تعداد

ویژگی‌ها:

استفاده از همه داده‌ها در محاسبه
اگر همه داده‌ها با یک عدد خاص جمع، تفریق، ضرب یا تقسیم شوند میانگین به همان نسبت
تغییر می‌کند.
مجموع انحراف داده‌ها از میانگین صفر است.

موارد استفاده:

برای داده‌های کمی

شاخص‌های مرکزی:

میان:

مقداری که نصف داده‌ها از آن کوچکتر باشند (مقدار میانی یا وسطی)

اگر داده‌ها را از کوچک به بزرگ مرتب نماییم، عدد m را میانه این داده‌ها می‌نامیم اگر نصف داده‌ها در سمت چپ و نصف داده در سمت راست این عدد قرار گیرد.

در مجموعه داده مرتب رتبه میانه برابر است با $(n+1)/2$

موارد استفاده:

داده‌های کمی یا کیفی ترتیبی

برتری نسبت به میانگین:

حساس نبودن به مقادیر بسیار بزرگ یا کوچک و متفاوت با سایر داده‌ها

قابلیت استفاده برای متغیرهای ترتیبی

شاخص‌های مرکزی:

نما:

داده‌ای که فراوانی آن نسبت به دیگر داده‌ها بیشتر باشد، (M)

ویژگی:

سادگی محاسبه

موارد استفاده:

برای همه‌ی انواع داده‌ها

شاخص‌های مرکزی:

چارک‌ها

چارک‌های اول، دوم و سوم (Q_1 ، Q_2 و Q_3 به ترتیب داده‌هایی هستند که ۰/۲۵، ۰/۵۰ و ۰/۷۵ داده‌ها از آن کوچک‌تر هستند. چارک داده‌ها را به ۴ قسمت مساوی تقسیم میکنند.

دهک‌ها

دهک‌های اول تا نهم (D_1 ، D_2 ، ...، D_9) به ترتیب داده‌هایی هستند که ۰/۱، ۰/۲، ...، ۰/۹ داده‌ها از آن کوچک‌تر هستند. دهک داده‌ها را به ۱۰ قسمت مساوی تقسیم میکند.

صدک‌ها

صدک‌های اول تا نود و نهم (P_1 ، P_2 ، ...، P_{99}) به ترتیب داده‌هایی هستند که ۰/۰۱، ۰/۰۲، ...، ۰/۹۹ داده‌ها از آن کوچک‌تر هستند. صدک داده‌ها را به ۱۰۰ قسمت مساوی تقسیم میکند.

چندک ها:

برای محاسبه چندک ها از فرمول کلی $(n+1)*p$ استفاده میشود.

مثال:

چارک اول $(n+1)/4$

چارک دوم $(n+1)/2$

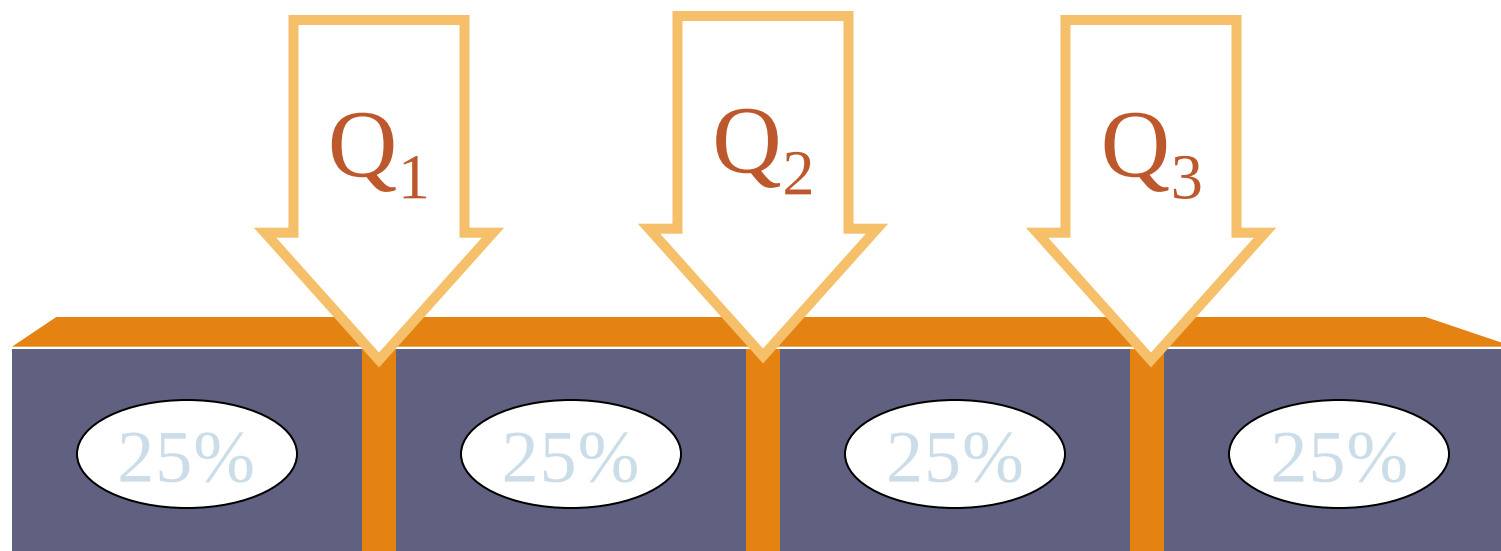
چارک سوم $3(n+1)/4$

دهک ۳ $3(n+1)/10$

دهک ۵ $5(n+1)/10$

صدک ۶۸ $68(n+1)/100$

چارک ها:



نمودار جعبه ای:

نمودار جعبه ای نموداری تصویری است که داده ها را بر اساس پنج مقدار نمایش می دهد. این مقادیر عبارتند از:

کوچک ترین داده

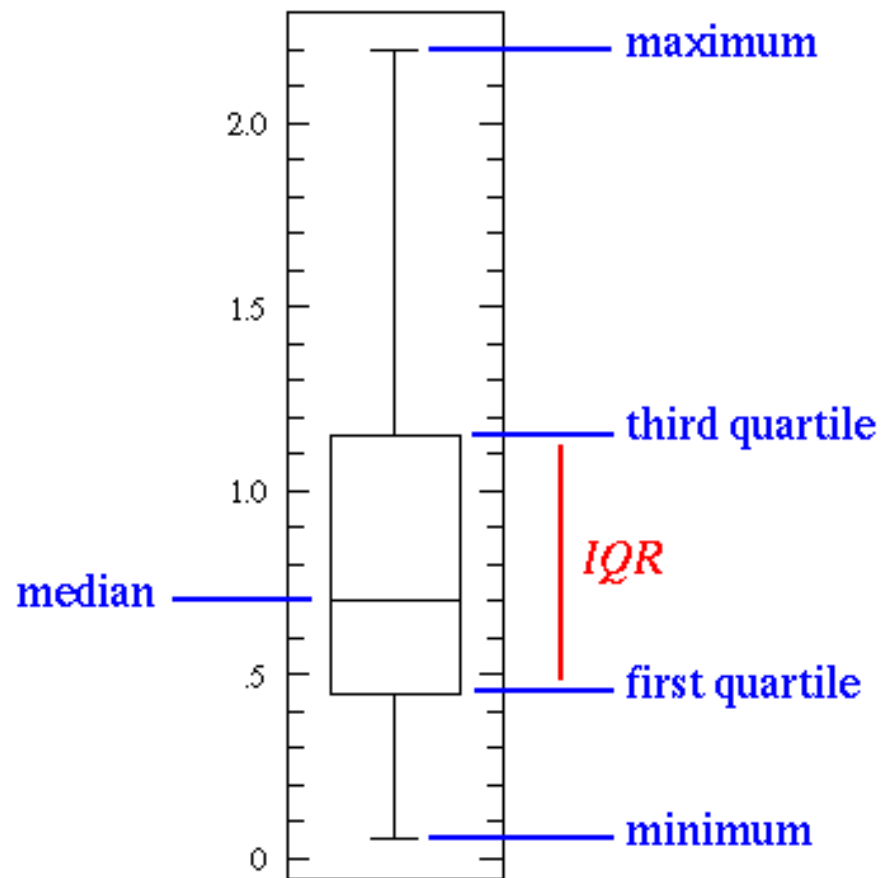
چارک اول

میانه

چارک سوم

بزرگ ترین داده

نمودار جعبه ای:



شاخص‌های پراکندگی:

اهمیت با ذکر مثال:

دو مجموعه داده:

۱۱، ۱۲، ۱۳، ۱۴، ۱۵

۶، ۷، ۱۳، ۱۹، ۲۰

میانگین و میانه برابر

اما تفاوت‌های ساختاری زیاد

لازم است علاوه بر اندازه‌گیری شاخص‌های مرکزی، در خصوص تغییرات، تفاوت‌ها و اختلاف‌های بین داده‌ها نیز کسب اطلاع نمود.

بدین منظور از شاخص‌های پراکندگی استفاده می‌شود.

شاخص‌های پراکندگی:

دامنه	کوچکترین $y_{(1)}$ و بزرگترین داده $y_{(n)}$	$R = y_{(n)} - y_{(1)}$
واریانس		$s^2 = \frac{1}{n} \sum_{i=1}^k f_i (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^k f_i x_i^2 - \bar{x}^2$
انحراف استاندارد		$s = \sqrt{\frac{1}{n} \sum_{i=1}^k f_i (x_i - \bar{x})^2}$
		$v = \frac{s}{\bar{x}}$

روشی ساده برای اندازه‌گیری پراکندگی داده‌ها

مطلوب نبودن در بسیاری از موارد تنها استفاده از دو داده

متوسط توان دوم اختلاف داده‌ها از میانگین. واحد اندازه‌گیری: توان دوم واحد اندازه‌گیری داده‌ها

شاخص پراکندگی بدون واحد مقایسه پراکندگی متغیرهای با واحد اندازه‌گیری متفاوت

جذر مثبت واریانس

واحد اندازه‌گیری مشابه داده‌ها

تغییر متغیر:

هدف ما کار با داده نرمال است. زمانی که داده ها نرمال نباشند تبدیل های برای نرمال کردن آنها اعمال میکنیم.

مهمترین تبدیلات :

توان ۲

جذر

لگاریتم

معکوس

تبدیلات باکس و کاکس