بسمه تعالی

موضوع : مبانی آمار و احتمال برای ورود به علم داده و تحلیل داده

بخش ششم:

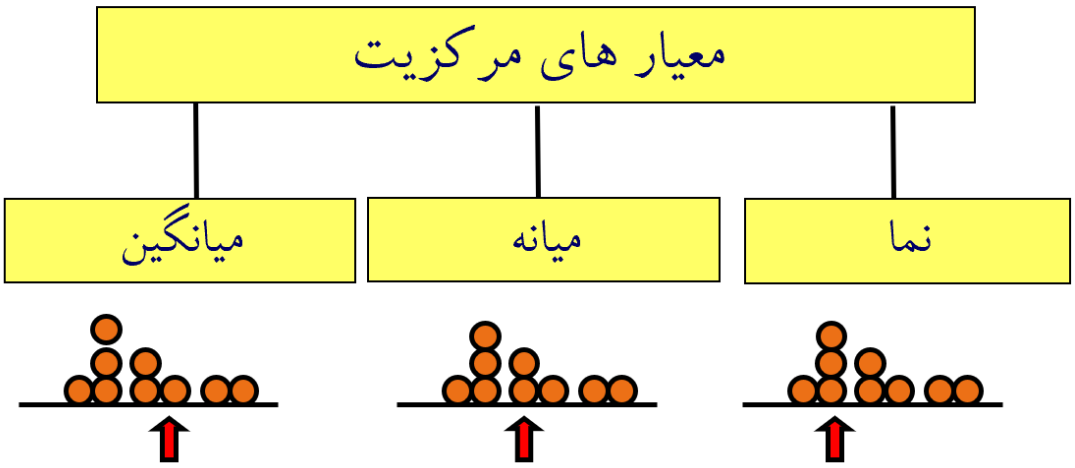
- **معیارهای مرکزی:**
  - میانگین(Mean)
  - میانه(Median)
  - نما(Mode)
  - مقایسه این سه شاخص
  - موارد استفاده هرکدام
  - مثال‌های دارای داده پرت(Outlier)

در این پارت از اموزش به بررسی **معیارهای تمرکز و مرکزیت داده ها** می پردازیم .

در اموزش های قبلی درباره آمار **توصیفی و جدول فراوانی صحبت** کردیم .

در حوزه آمار توصیفی برای بررسی دقیق تر رفتار داده ها نیاز به شاخصی که بررسی کند که مرکز داده ها یا نقطه تعادل داده ها کجاست. به بیان ساده، این شاخص‌ها تلاش می‌کنند یک عدد را به‌عنوان نماینده کل داده‌ها معرفی کنند.

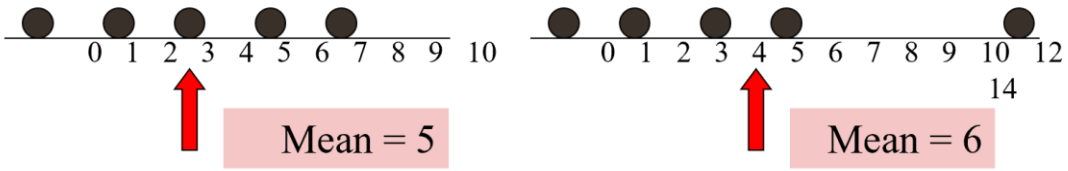
سه آماره (**چون داده ها از روی داده های نمونه گیری و ثبت شده بدست آمده/آماره، هر عددی است که از روی نمونه آماری محاسبه شود.**) اصلی در این بخش :



میانگین(Mean) : ابتدا با شناخته‌شده‌ترین معیار مرکزی یعنی میانگین آشنا می‌شویم.

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N} = \frac{X_1 + X_2 + \cdots + X_N}{N}$$



- میانگین متداولترین معیار تمرکز داده ها می باشد.
- نقاط دور افتاده روی آن تاثیر می گذارند.
- محدودیت میانگین: به داده‌های بسیار بزرگ یا بسیار کوچک، یعنی داده‌های پرت (Outlier)، حساس است.

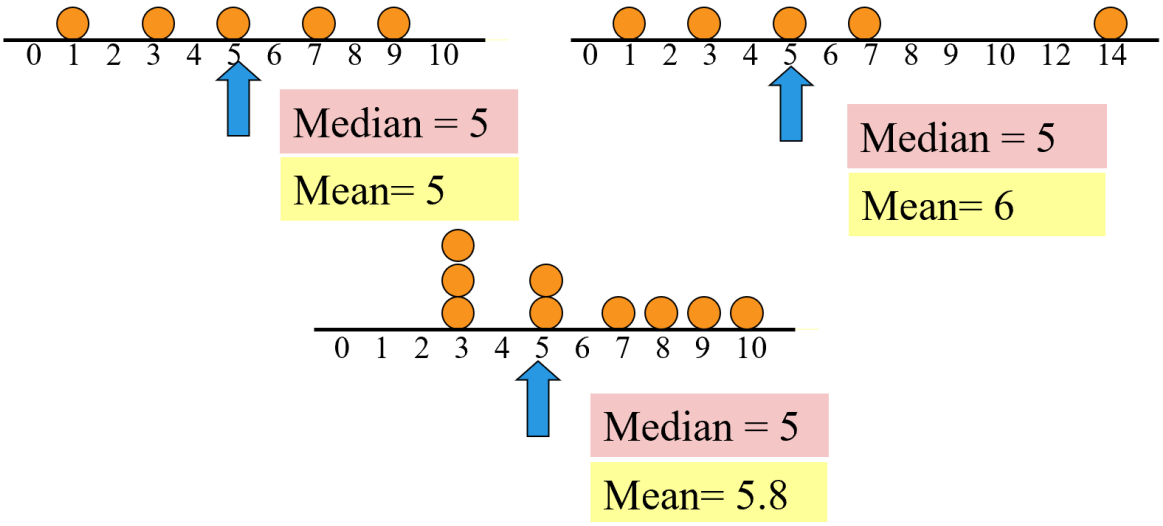
میانه(Median) **:**

میانه، عددی است که دقیقاً در وسط داده‌های مرتب‌شده قرار می‌گیرد.

برای محاسبه میانه ابتدا داده‌ها را از کوچک به بزرگ مرتب می‌کنیم.

اگر تعداد داده‌ها فرد باشد، عدد وسط میانه است.

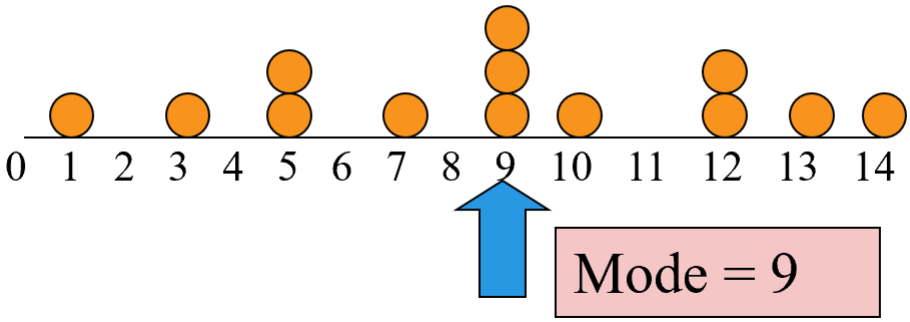
اگر تعداد داده‌ها زوج باشد، میانگین دو عدد وسط را محاسبه می‌کنیم.



مزیت میانه: برخلاف میانگین، میانه تحت تأثیر داده‌های پرت قرار نمی‌گیرد.

نما(Mode) **:** نما مقداری است که بیشترین تعداد تکرار را در داده‌ها دارد.

این شاخص تنها معیار مرکزی است که برای داده‌های کیفی نیز قابل استفاده است.



نکته !!! نما تحت تاثیر نقاط دور افتاده قرار نمی گیرد.

نکته !!! داده های نمونه ممکن است چند نما داشته باشند.

مقایسه سه آماره:

ویژگی	میانگین	میانه	نما
از تمام داده‌ها استفاده می‌کند	✅	❌	❌
به داده‌های پرت حساس است	✅	❌	❌
برای داده‌های کیفی قابل استفاده است	❌	❌	✅
محاسبه آسان	✅	✅	✅