

باسمه تعالی

نکات " دامنه بین چارکی، صدک‌ها، و نمودار جعبه‌ای "

صدک‌ها: (Percentiles)

- صدک‌ها مقادیر داده را به 100 بخش مساوی تقسیم می‌کنند و نشان می‌دهند که یک مقدار مشخص در کجای مجموعه داده قرار دارد.
- مثال: صدک 90 نشان‌دهنده مقداری است که 90 درصد داده‌ها کمتر از آن هستند.

دامنه بین چارکی: (IQR)

- دامنه بین چارکی برابر است با فاصله بین چارک اول (Q1) و چارک سوم (Q3).
- IQR نشان‌دهنده پراکندگی میانه 50% داده‌ها است و به شناسایی داده‌های پرت کمک می‌کند.
- فرمول $IQR = Q3 - Q1$:

نمودار جعبه‌ای: (Boxplot)

- یک نمایش گرافیکی برای توزیع داده است که چارک اول (Q1)، چارک سوم (Q3)، میانه (Median)، و داده‌های پرت را نشان می‌دهد.
- بخش‌های اصلی:
 - **خط میانی:** نشان‌دهنده میانه داده‌ها است.
 - **جعبه:** محدوده بین Q1 و Q3 را نشان می‌دهد. (IQR)
 - **ویسکرها:** کوچک‌ترین و بزرگ‌ترین مقادیر غیرپرت را نمایش می‌دهند.
 - **نقاط پرت:** داده‌هایی که خارج از محدوده $Q1 - 1.5 \times IQR$ تا $Q3 + 1.5 \times IQR$ هستند.

کاربردها در: AI/ML

- **شناسایی داده‌های پرت:** برای بهبود عملکرد مدل‌ها.
- **مهندسی ویژگی:** تحلیل پراکندگی داده‌ها برای نیاز به نرمال‌سازی یا استانداردسازی.
- **پیش‌پردازش داده:** تصمیم‌گیری درباره نحوه آماده‌سازی داده‌ها.