آمار و احتمالات مهندسی

جلسه سوم

شاخص های آماری

مهمترین شاخص ها

- شاخص های مرکزی (Central Tendency Indexes)
- شاخص های پراکندگی (Dispersion Indexes)
- شاخص های توزیع یا تقارن (Distribution Indexes)

شاخص های آماری

شاخص های مرکزی

- میانگین (Mean)
  - میانگین حسابی (Arithmetic mean)
  - میانگین اصلاح شده (Trimmed mean)
  - میانگین وزنی (Weighted mean)
  - میانگین هندسی (Geometric mean)
  - میانگین همساز یا توافقی (Harmonic mean)
- میانه (Median)
- نما (Mode)
- چندک ها (Quintile)
  - چارک (Quartiles)
  - دهک ها (Deci les)
  - صدک ها (Percentiles)

شاخص های آماری

مهمترین شاخص ها

شاخص های مرکزی (Central Tendency Indexes)

شاخص های پراکندگی (Dispersion Indexes)

شاخص های توزیع یا تقارن (Distribution Indexes)

شاخص های آماری

شاخص های پراکندگی

معیاری که میزان پراکندگی داده ها را نشان دهد.

- برای داده های کیفی، همان جدول فراوانی میتواند به عنوان شاخص یا معیار پراکندگی در نظر گرفته می شود.
- برای داده های کمی، مقداری است که بزرگی آن نشان دهنده دوری مقادیر از یکدیگر است.
  - برد یا دامنهٔ تغییرات (Range) یک شاخص پراکندگی است
  - برد میان چارکی- Interquartile Range
  - واریانس- Variance
  - انحراف معیار- Standard Deviation

شاخص های آماری

شاخص های پراکندگی

برد یا دامنهٔ تغییرات (Range):

برای داده های کمی کاربرد داشته و فاصله بین بزرگترین مقدار و کوچکترین مقدار را بیان می کند. و حداکثر میزان پراکندگی را نشان می دهد.

$$R = X_{max} - X_{min}$$

محدودیت شاخص برد این است که قادر به نمایش پراکندگی نسبت به یک نقطهٔ تمرکز نیست و فقط بزرگترین و کوچکترین داده در آن نقش داشته و بقیه داده ها تاثیر ندارند.

برد میان چارکی (Interquartile Range):

فاصله بین چارک سوم (Q_3) و چارک اول (Q_1) را بیان می کند.

$$Q = Q_3 - Q_1$$

مثال: برای داده های بین ۰ تا ۲۰ چارک اول برابر با **۴/۵** و چارک سوم برابر با **۱۵/۵** است.

بنابراین برد میان چارکی برابر با **۱۱** خواهد بود.

$q = (n+1) * p$ ➔ $q = (22) * .25 = 5.5$

چون ۵/۵ رند نیست به دو قسمت صحیح (r) و اعشاری (w) تبدیل می شود و از رابطه زیر داریم:

$$Q_1 = (1-w)X_r + w(X_{r+1}) = 0.5 * 4 + 0.5 * 5 = 4.5$$

شاخص های آماری

شاخص های پراکندگی

واریانس (Variance):

برای داده های کمّی کاربرد داشته و برابر با متوسط مربع فاصله مقادیر از میانگین آنهاست.

واحد اندازه گیری به صورت مربع (توان ۲) در می آید.

$$Var(X) = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}$$

برای مثال واریانس داده های از ۰ تا ۲۰ برابر با ۳۸/۶۶۷ است.

نکته ۱: اگر مقادیر یک سری داده را با یک عدد ثابت مثل b جمع کنیم، واریانس تغییری نخواهد کرد.

نکته ۲: اگر این سری را در مقدار ثابت a ضرب کنیم، واریانس به صورت حاصلضرب واریانس اولیه در مربع مقدار a در خواهد آمد.

انحراف معیار (Standard Deviation):

اگر از واریانس جذر بگیریم مقدار انحراف از معیار بدست می آید و واحد آن همان واحد داده ها است.

$$SD(X) = \sqrt{Var(X)}$$

انحراف از معیار داده های از ۰ تا ۲۰ برابر با ۶/۰۵۵ است.

نکته: اگر داده های یک سری را در مقدار ثابت a ضرب کنیم، مقدار انحراف معیار نیز به صورت حاصلضرب مقدار a در انحراف از معیار اولیه در خواهد آمد.

شاخص های آماری

مهمترین شاخص ها

- شاخص های مرکزی (Central Tendency Indexes) ✓
- شاخص های پراکندگی (Dispersion Indexes) ✓
- شاخص های توزیع یا تقارن (Distribution Indexes)

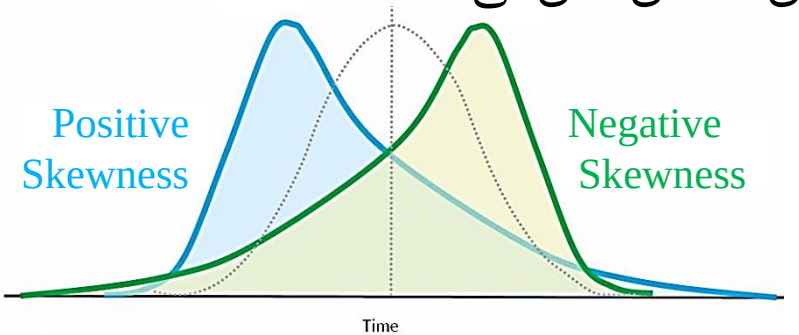
شاخص های آماری

شاخص های تقارن

اگر جامعه ایده آل را جامعه ای با توزیع نرمال استاندارد در نظر بگیریم، میزان عدم تقارن هر جامعه نسبت به این توزیع نرمال در شاخص های تقارن مورد بررسی قرار می گیرد.

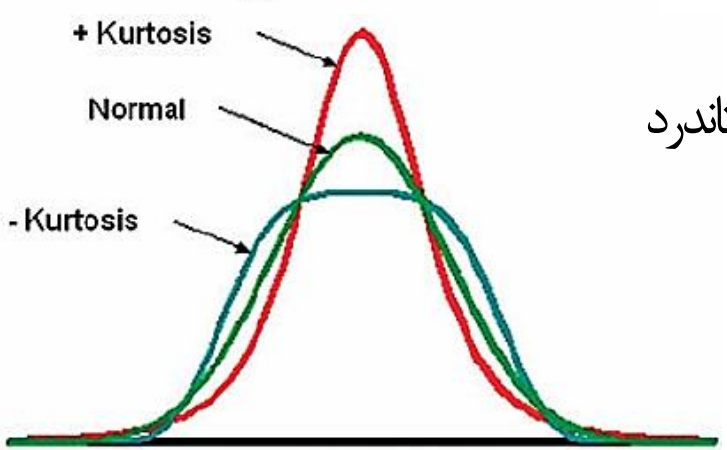
- شاخص چولگی (Skewness)

میزان عدم تقارن داده ها نسبت به توزیع نرمال استانداد را با این شاخص نشان می دهند.



- شاخص کشیدگی یا برجستگی (Kurtosis)

میزان کشیدگی یا پخی داده ها نسبت به توزیع نرمال استاندرد



شاخص های آماری

شاخص های تقارن

برای محاسبه **چولگی** (Skewness) معمولاً از سه نوع ضریب چولگی استفاده می شود.

1) ضریب چولگی اول پیرسون (b_1) که در آن M نما و S انحراف معیار است.

$$b_1 = \frac{(\overline{X} - M)}{S}$$

2) ضریب چولگی دوم پیرسون (b_2) که در آن m میانه است.

$$b_2 = \frac{3(\overline{X} - m)}{S}$$

3) ضریب گشتاوری چولگی (g) که در آن m_3 گشتاور مرکزی مرتبه سوم است.

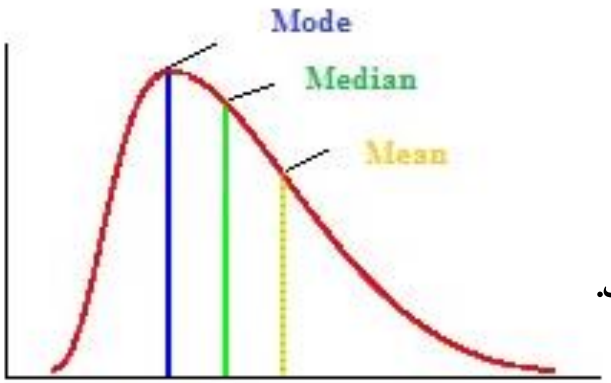
$$g = \frac{m_3}{s^3} \qquad m_r = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^r}{n}$$

شاخص های آماری

شاخص های تقارن

ضریب چولگی: مقدار این ضریب برای منحنی نرمال استاندارد برابر با صفر است.

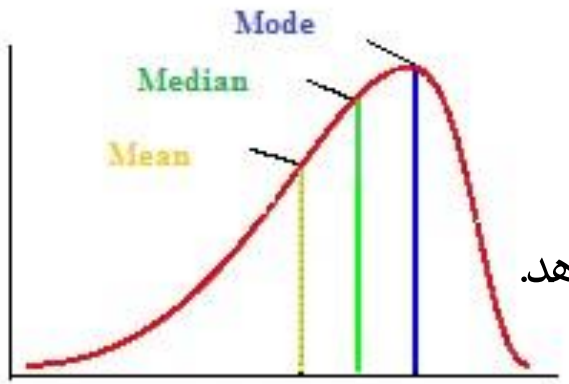
Right Skewness یا چولگی به راست:



Right-Skewed (Positive Skewness)

- ضریب چولگی مثبت است
- تمایل منحنی داده ها به سمت راست است
- به دلیل وجود مقادیر بزرگ و بسیار بزرگ نسبت به سایر داده ها رخ می دهد.

Left Skewness یا چولگی به چپ:



Left-Skewed (Negative Skewness)

- ضریب چولگی منفی است
- تمایل منحنی داده ها به سمت چپ است
- به دلیل وجود مقادیر کوچک و بسیار کوچک نسبت به سایر داده ها رخ می دهد.

شاخص های آماری

شاخص های تقارن

مثال ضریب چولگی:

طول عمر ۱۰ باتری دارای میانگین ۳/۵ سال و میانه ۳/۴۸ سال و انحراف استاندارد ۱/۶۵ سال است. ضریب چولگی دوم پیرسن برابر است با؟

$$b_2 = \frac{3\times(3.5-3.48)}{1.65} = 0.036$$

مقدار چولگی کوچک و مثبت بدست آمد. پس منحنی کمی چوله به راست است.

نکته: با ضرب مقادیر داده ها در یک مقدار ثابت مثل a ضریب چولگی تفاوت نخواهد کرد.

شاخص های آماری

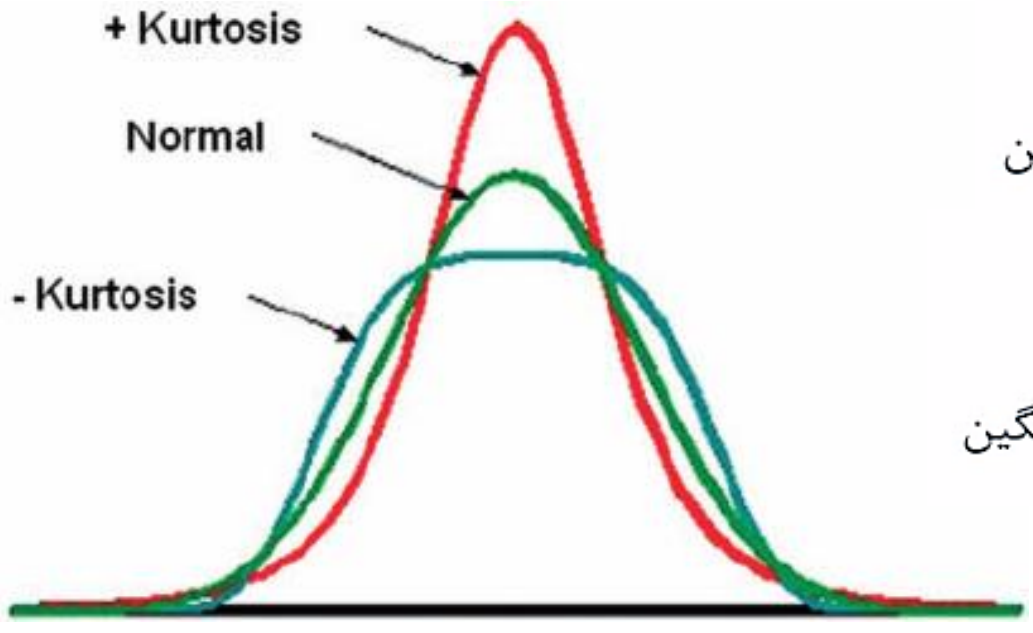
شاخص های تقارن

مقدار شاخص برجستگی یا کشیدگی (Kurtosis) برای منحنی نرمال استاندارد برابر با صفر است.

$$k = \frac{m_4}{S^4} - 3$$

m_4 گشتاور مرکزی مرتبه چهارم

s^4 انحراف استاندارد به توان ۴



- **کشیدگی**
- ضریب کشیدگی مثبت
- تمرکز بیش از حد مقادیر حول میانگین

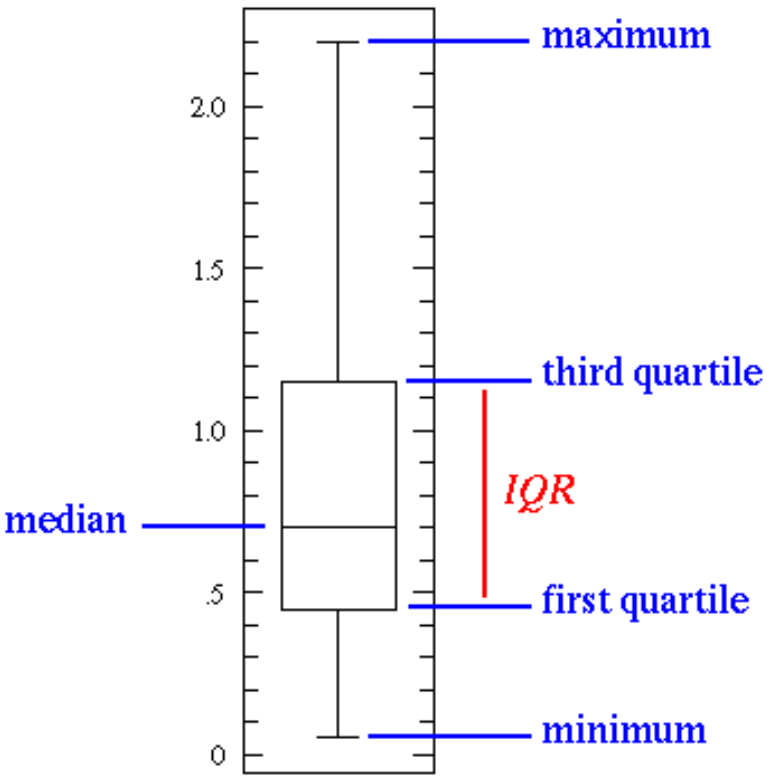
- **پخی**
- ضریب کشیدگی منفی
- پراکندگی بیش از حد مقادیر حول میانگین

در مسئله قبل گشتاور مرکزی مرتبه چهارم برای طول عمر ۱۰۰ باتری برابر با ۹.۲۳ است. ضریب برجستگی برابر با ۱.۷۵- خواهد بود، که نشان‌دهنده پخی در منحنی فراوانی نسبت به منحنی نرمال استاندارد است.

نمودار جعبه ای (Box-Plot)

- ✓ برای مقایسه دو یا چند جامعه آماری در کنار یکدیگر از این نمودار استفاده می شود.
- ✓ در این نمودارد چارکها، حداقل و حداکثر مشاهدات و میانه همزمان قابل روئیت است.
- ✓ از این نمودار تقارن یا عدم تقارن داده ها و همچنین وجود داده های پرت قابل مشاهده است

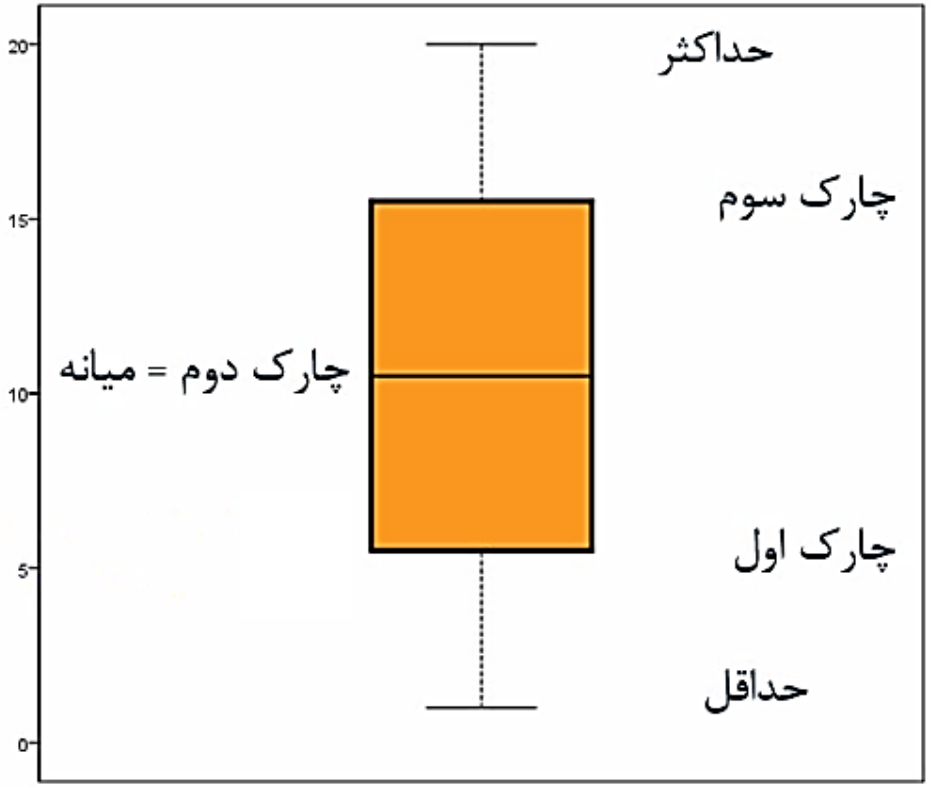
(دادهٔ **پرت** مقداری است که یک و نیم برابر دامنهٔ میان چارکی (Q_3-Q_1) از Q_1 و Q_3 فاصله داشته باشد.)



روش رسم:

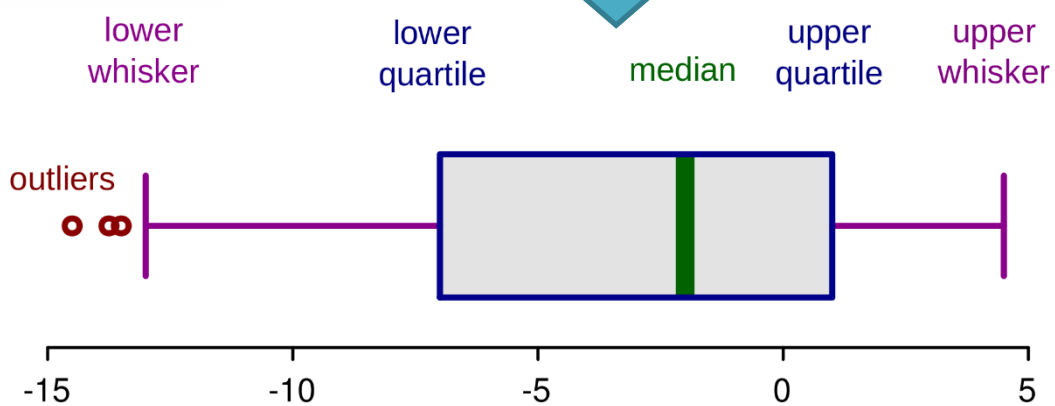
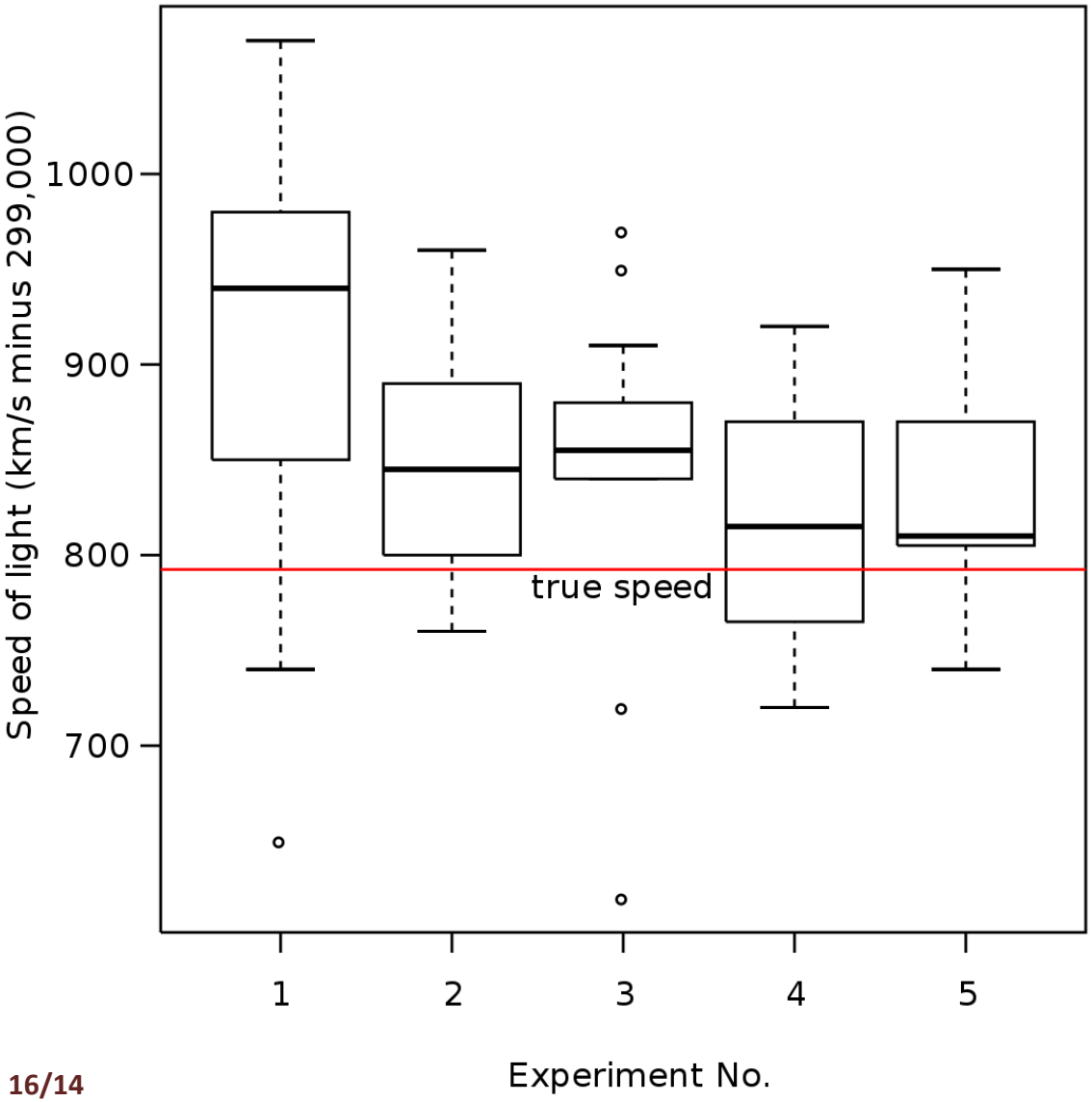
1) حداقل و حداکثر داده ها پیدا شود.
2) چارک های اول، دوم و سوم پیدا شوند.
3) مستطیلی به طول فاصله چارک اول تا سوم رسم شود و محل میانه در داخل مستطیل با یک خط افقی نشان داده شود.
4) مقدار حداقل و حداکثر در دو طرف مستطیل به کمک دو خط نشان داده شده و با خط عمود به مستطیل وصل شود.

نمودار جعبه ای (Box-Plot)



نمودار جعبه برای دنباله اعداد از ۱ تا ۲۰

نمونه ای دیگر برای Box Plot



نحوه رسم نمودار جعبه ای (Box-Plot) در اکسل