آمار و احتمالات مهندسی
اندازه پراکندگی و چولگی

دکتر بهناز بیگدلی
دانشکده مهندسی عمران
دانشگاه صنعتی شاهرود

ساختار درس

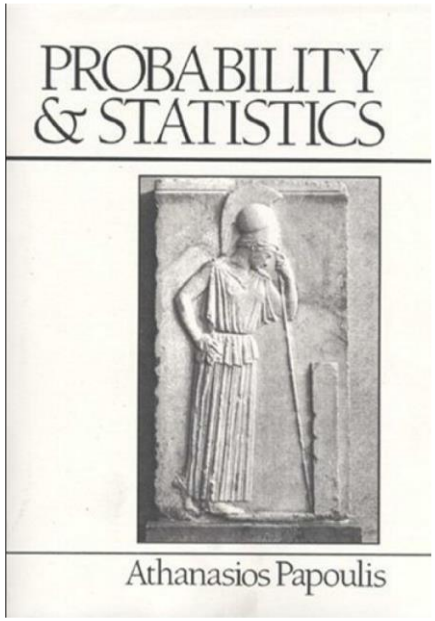
نحوه ارزیابی

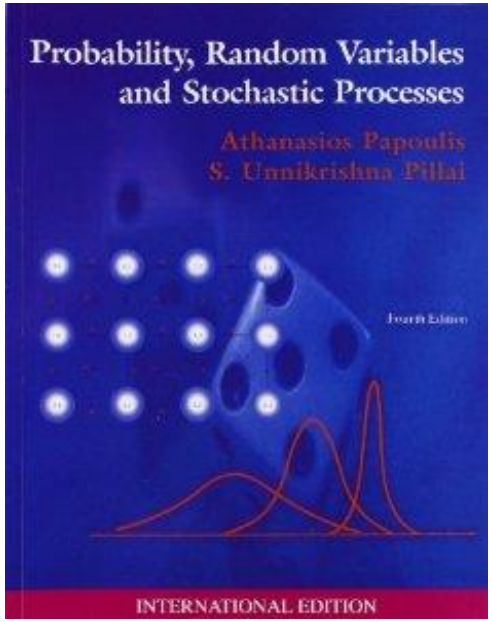
امتحان پایان ترم (۱۴)

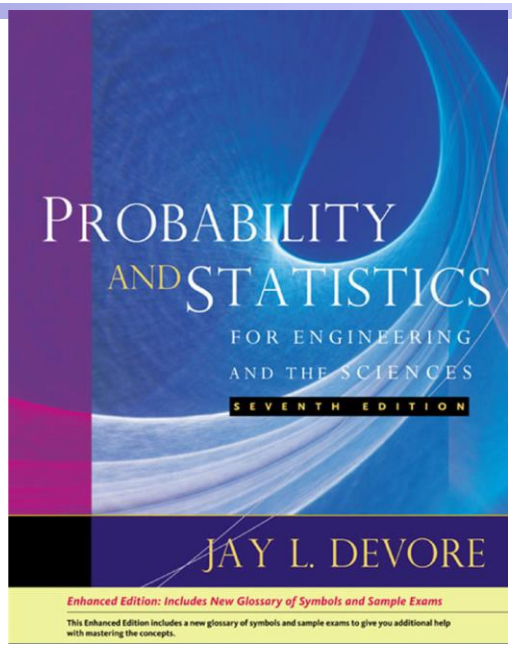
امتحان میان ترم (۴)

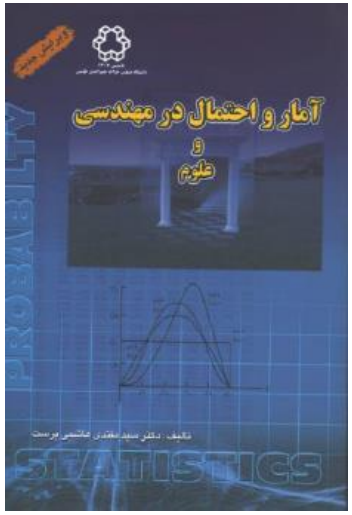
تمرینات و کوئیزها(۲)

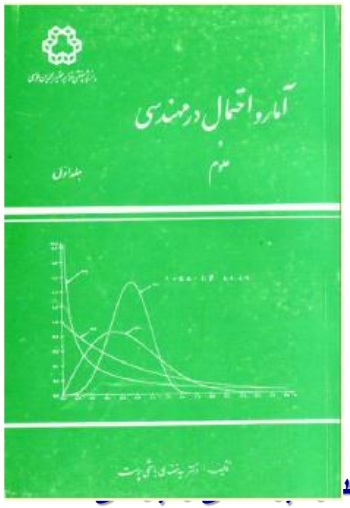
منابع


PROBABILITY
& STATISTICS
Athanasios Papoulis


Probability, Random Variables
and Stochastic Processes
Athanasios Papoulis
S. Unnikrishna Pillai
Fourth Edition
INTERNATIONAL EDITION


PROBABILITY
AND STATISTICS
FOR ENGINEERING
AND THE SCIENCES
SEVENTH EDITION
JAY L. DEVORE
Enhanced Edition: Includes New Glossary of Symbols and Sample Exams
This Enhanced Edition includes a new glossary of symbols and sample exams to give you additional help with mastering the concepts.


آمار و احتمال در مهندسی
و
علوم


آمار و احتمال در مهندسی

فهرست و عناوین درس

I. بخش اول: آمار مقدماتی (توصیفی)

1. آمار مقدماتی
2. اندازه‌گیری پراکندگی، گشاورها، چولگی و کشیدگی
3. برازش خط و منحنی بر داده‌ها
4. توزیع‌های دو بعدی و ضریب همبستگی

II. احتمال

1. احتمال و فضای نمونه
2. فضای نمونه با عناصر متعدد
3. احتمالات شرطی و نابستگی
4. تابع چگالی احتمال، تابع توزیع و امید ریاضی
5. توزیع‌های گسسته
6. توزیع‌های پیوسته مهم
7. نظریه برآورد
8. آزمون‌های فرض

اندازه‌گیری پراکندگی،
گشاورها، چولگی و کشیدگی

شاخص های پراکندگی:

۱- دامنه، فاصله دسته، مراکز دسته

۲- میانگین قدر مطلق انحرافات

۳-واریانس

۴-انحراف معیار

۵-متغیرهای استاندارد

۶-ضریب تغییر یا تعیین

۷-انحراف چارکی

۸-گشتاورها

۱ – دامنه، فاصله دسته، مراکز دسته

$$X_1, X_2, \ldots, X_i, \ldots, X_n$$

$$f_1, f_2, \ldots, f_i, \ldots, f_n$$

$$N = \sum_{i=1}^{N} f_i$$

$$R = X_{MAX} - X_{MIN}$$

$$c = \frac{R+1}{n}$$

$$c = \frac{R}{1 + 3.3 \times \log(N)}$$

$$k = ۱ + ۳/۳۲ \log n$$

R دامنه داده‌ها

n تعداد دسته‌ها

N جمع کل فراوانی‌ها

c فاصله دسته‌ها

K تعداد دسته ها

۱- انحراف میانگین از عدد دلخواه β

$$x_1, x_2, \ldots, x_i, \ldots, x_n$$

$$f_1, f_2, \ldots, f_i, \ldots, f_n$$

$$\frac{1}{N}\sum_{i=1}^{N}|x_i - \beta| = \frac{1}{N}\sum_{i=1}^{N} f_i|x_i - \beta|$$

۲- میانگین قدر مطلق انحرافات

عبارت $\sum f_i(x_i - \mu)$ را می‌توان به عنوان شاخص پراکندگی در نظر گرفت ولی از آن جا که مقدار آن همواره صفر است، این مقدار مفید نیست لذا این شاخص را می توان به صورت

$$AD = MAD = \frac{1}{N}\sum f_i|x_i - \mu|$$

تعریف کرده و آن را به عنوان معیاری برای پراکندگی در نظر گرفت. عیبی که بر این شاخص وارد است آن است که در بعضی از مواقع استفاده از قدرمطلق کار مشکلی است و در ادامه نشان خواهیم داد که به این دلیل شاخص دیگری به نام واریانس به وجود آمد.

۳-واریانس

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$S^2 = \frac{1}{n}\sum_{i=1}^{n} f_i(x_i - \bar{x})^2$$

ویژگی های واریانس جامعه:

۱-واریانس عدد ثابت C برابر با صفر است.

۲-اگرمقدار ثابت a رابه مشاهدات اضافه یا ازآنها کم کنیم واریانس تغییر نمی‌کند.

۳-اگر مشاهدات در مقدار ثابت K ضرب یا برآن تقسیم شود واریانس جدید از ضرب یا تقسیم واریانس قدیم در K^2 بدست می‌آید.

۴-انحراف معیار

انحراف معیار در نمونه جذر واریانس یا پراش می باشد.

$$\sqrt{S^2} = S = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

μ= میانگین جامعه

δ^2 = واریانس جامعه

و جذر آن انحراف معیار جامعه

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i - \mu)^2$$

خواص واریانس و انحراف معیار

۱) واریانس هر عدد ثابت صفر است.

$$Var(c) = ۰$$

۲) جمع و تفریق اثری در واریانس و انحراف معیار ندارد. ولی اگر تک‌تک مشاهدات را در عدد ثابتی مانند a ضرب یا بر عدد ثابتی مانند a تقسیم کنیم، واریانس مشاهدات در مجذور آن عدد ضرب می‌شود. و انحراف معیار در قدر مطلق عدد a ضرب یا تقسیم می‌شود.

$$\begin{cases} Var(aX \pm b) = a^{۲}.Var(x) \\ \sigma(ax \pm b) = |a|\,\sigma_x \end{cases}$$

۳) اگر داده‌های $X_1, X_2, \ldots, X_n$ با یکدیگر برابر باشند واریانس و انحراف معیار بین آن ها صفر است و برعکس اگر واریانس یا انحراف معیار داده هایی صفر باشد آن داده‌ها با یکدیگر برابرند.

$$x_۱ = x_۲ = \cdots = x_n \Leftrightarrow \begin{cases} \sigma^{۲}(x_۱, x_۲, \ldots, x_n) = ۰ \\ \sigma(x_۱, x_۲, \ldots . x_n) = ۰ \end{cases}$$

تصحیح شپارد

در داده‌های طبقه‌بندی شده (پیوسته) برای محاسبه میانگین و واریانس از نماینده طبقات (x_i) استفاده می‌شود. در میانگین به علت مثبت و منفی بودن تفاضل‌ها این خطا جبران می‌شود ولی در واریانس این خطاها به توان دو رسیده و مقدار به دست آمده از مقدار واریانس واقعی بیشتر می‌شود. برای بر طرف کردن این مشکل، رابطه زیر توسط شپارد معرفی شد که به تصحیح شپارد معروف است.

$$\sigma_c^2 = \sigma_x^2 - \frac{C^2}{12}$$

C^2 مربع فاصله طبقات می‌باشد.

توجه کنید تصحیح شپارد در مواردی بکار می‌رود که متغیر پیوسته باشد و تعداد دست کم هزار باشد و همچنین توزیع فراوانی از نوع متقارن یا اندکی متقارن باشد.

نیمه واریانس (واریانس داده‌های نامطلوب)

یکی دیگر از شاخص‌های پراکندگی، نیمه واریانس است که برای انحرافات نامطلوب و نامناسب بکار می‌رود. در داده‌های مربوط به سود، مقادیر کوچکتر از میانگین و در داده‌های مربوط به زیان داده‌های بزرگتر از میانگین، نامطلوب هستند. این شاخص از رابطه زیر به دست می‌آید.

$$S.V = \frac{\sum_{i=1}^{k}(x_i - \mu_x)^2}{K}$$

در اینجا فرض بر این است که تعداد مشاهدات جامعه N تاست و فقط K تا (K<N) از آن ها نامطلوب است.

۵-متغیرهای استاندارد

$$z_i = \frac{x_i - \bar{x}}{S} \quad , \qquad i = 1,2,\ldots,n$$

ویژگی های متغیرهای استاندارد:

۱- میانگین متغیرهای استاندارد برابر صفر است.

$$\sum_{i=1}^{n} z_i = \sum_{i=1}^{n} \frac{x_i - \bar{x}}{S} = \frac{1}{S}\sum_{i=1}^{n}(x_i - \bar{x}) = \frac{1}{S}(\sum x_i - n\bar{x}) = \frac{1}{S}(n\bar{x} - n\bar{x}) = 0$$

۲-واریانس متغیرهای استاندارد برابر با ۱ است .

۳- متغیرهای استاندارد فاقد واحد اندازه گیری هستند.

۴- مقدار Z_i می تواند، منفی، صفر یا مثبت باشد.

۶-ضریب تغییر و ضریب انحراف چندک

$$C.V = \frac{S}{\bar{x}} = \frac{\sqrt{\frac{1}{n}\sum(x_i - \bar{x})^2}}{\frac{1}{n}\sum_{i=1}^{n} x_i} = \frac{\sqrt{n\sum(x_i - \bar{x})^2}}{\sum x_i}$$

ویژگی‌های ضریب تغییر

۱- به واحد اندازه‌گیری بستگی ندارد.

۲- برای مقایسه دو صفت از یک جامعه با واحدهای اندازه‌گیری متفاوت مورد استفاده قرار می‌گیرد.

۳- مجموعه مشاهداتی که دارای C.V کمتری است از سازگاری و همگنی بیشتری برخوردار هستند.

۴- اگر به تک تک داده‌ها مقدار مثبتی اضافه کنیم ضریب تغییرات کوچکتر می‌شود و اگر از تک تک داده‌ها مقدار مثبتی را کم کنیم ضریب تغییرات بزرگتر می‌شود.

۵- اگر داده‌ها را در عددی مثبت ضرب کنیم ضریب تغییرات تغییری نمی‌کند ولی اگر داده‌ها را در عددی منفی ضرب کنیم، ضریب تغییرات قرینه می‌شود.

ضریب انحراف چندک

$$\frac{Q_3 - Q_1}{Q_3 + Q_1}$$

۷-دامنه و انحراف چارکی

$$IQR = Q_3 - Q_1$$

دامنه چارکی

$$SIQR = \frac{Q_3 - Q_1}{2}$$

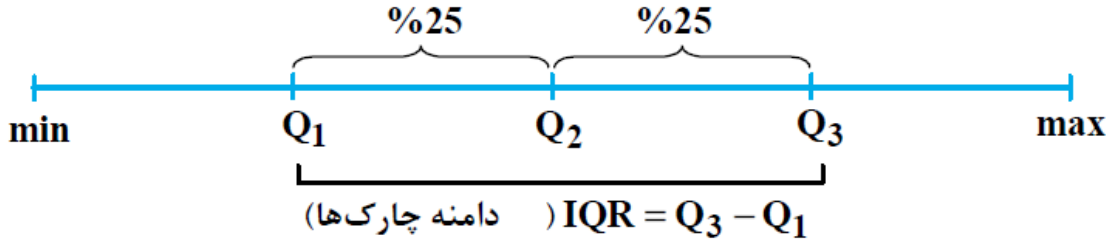
انحراف چارکی

ویژگیهای انحراف چارکی:

۱- این شاخص چون میزان پراکندگی در اطراف مرکز توزیع را نشان می‌دهد از شاخص دامنه با ثبات تر است.

۲- این شاخص چون شامل ۲۵٪ از مشاهدات کوچک و بزرگ نیست تحت تأثیر داده‌های پرت قرار نمی‌گیرد.

۳- این شاخص برای داده‌های کلاس بندی نیز قابل محاسبه است.



نکته:

در توزیع‌های نرمال، رابطه بین انحراف معیار σ ، انحراف چندک SIQR و میانگین قدر مطلق انحراف از میانگین MAD

$$SIQR = \frac{2}{3}\sigma$$

$$MAD = \frac{4}{5}\sigma$$

۸-گشتاورها

$$\mu'_r = \frac{1}{n}\sum_{i=1}^{n}(x_i - a)^r \qquad , \qquad r = 1,2,\ldots$$

ویژگی‌های گشتاورهای مرکزی:

۱- $\mu_1 = 0$, $r=1$

۲- $r=2$

$$\mu_2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = S^2$$

۳- تغییر در مبدأ یا اضافه و کم کردن مقدار ثابت به مشاهدات تغییری در μ_r ندارد

۴-باتغییر در مقیاس یا ضرب و تقسیم کردن مقدار ثابت در مشاهدات، μ_r در توان r ام مقدار ثابت ضرب یا تقسیم می شود

۵- $$\mu_2 = \mu'_2 - (\bar{x} - a)^2$$

۸-گشتاورها

$$\mu_r = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^r$$

اگر $\bar{x} - a = d$ خواهیم داشت:

$$\mu_r = (\mu'_r - d)^r$$

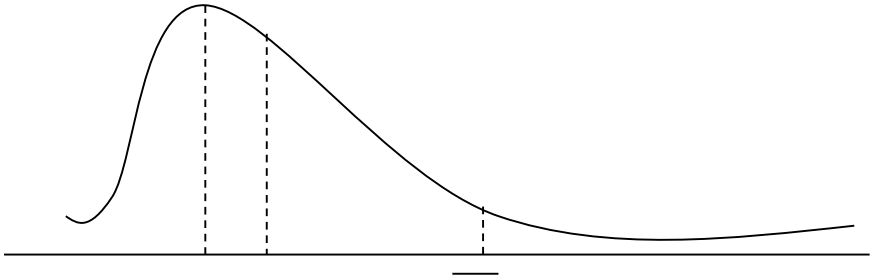
$$\mu'_r = (\mu_r + d)^r$$

چولگی

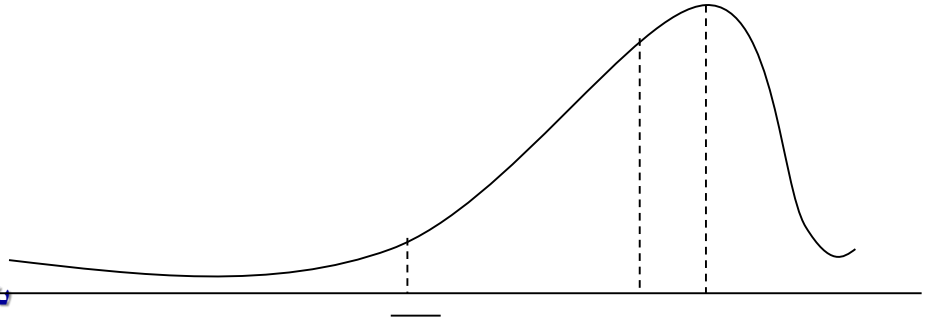
چولگی عددی است که برای عدم وجود تقارن بکار می‌رود.

انواع چولگی عبارتند از:

۱- **چولگی مثبت** : منحنی فراوانی در سمت راست دامنه داده‌ها قرار گیرد.



۲- **چولگی منفی**: منحنی فراوانی در سمت چپ دامنه داده‌ها قرار گیرد.



چولگی

معیارهای محاسبه میزان چولگی عبارتند از:

۱- ضریب چولگی کارل - پیرسن

$$Sk_1 = \frac{\bar{x} - Mo}{\sigma}$$

$$Sk_2 = \frac{3(\bar{x} - Md)}{\sigma}$$

۲- ضریب چولگی بولی (چارکی)

$$Sk_Q = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$$

۳- ضریب چولگی صدکی

$$Sk_P = \frac{P_{90} + P_{10} - 2P_{50}}{P_{90} - P_{10}}$$

چولگی

تفسیر ضریب چولگی

۱- اگر $SK > 0$ باشد چولگی به سمت راست است.

۲- اگر $SK < 0$ باشد چولگی به سمت چپ است.

۳- اگر $SK = 0$ باشد توزیع متقارن است.

۴- اگر $|SK| \leq 0.1$ باشد، چولگی خفیف است و جامعه از نظر تقارن تقریباً متقارن (نرمال) است.

۵- اگر $0.1 < |SK| \leq 0.5$ چولگی کم است و تفاوت اندکی با توزیع نرمال وجود دارد.

۶- اگر $|SK| > 0.5$ باشد چولگی زیاد است و تفاوت با توزیع نرمال فاحش است.

چولگی

معیارهای محاسبه میزان چولگی عبارتند از:

۴- ضریب چولگی پیرسن β

(ضریب کشیدگی)

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} \qquad \beta_2 = \frac{\mu_4}{\mu_2^2}$$

۵- ضریب چولگی پیرسن γ

$$\gamma_1 = +\sqrt{\beta_1} \qquad \gamma_2 = \beta_2 - 3$$

۶- ضریب کشیدگی چندکی

$$E_P = \frac{SIQR}{P_{90} - P_{10}} - 0.263$$

$$SIQR = \frac{Q_3 - Q_1}{2}$$

چولگی

تفسیر ضریب کشیدگی

۱- اگر $\gamma_2 > 0$ باشد، توزیع از توزیع نرمال کشیده‌تر و بلندتر است.

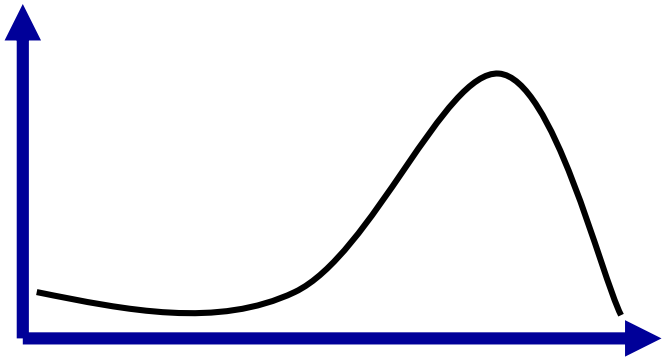
۲- اگر $\gamma_2 < 0$ باشد، توزیع از توزیع نرمال، کوتاه‌تر است.

۳- اگر $|\gamma_2| \leq 0.1$ باشد شکل توزیع تقریباً متقارن (نرمال) است.

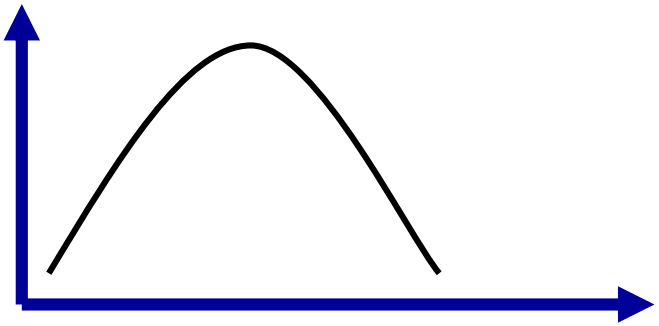
۴- اگر $0.1 < |\gamma_2| \leq 0.5$ باشد، شکل توزیع دارای تفاوت اندکی با نرمال است.

۵- اگر $|\gamma_2| > 0.5$ باشد، تفاوت با توزیع نرمال فاحش است.

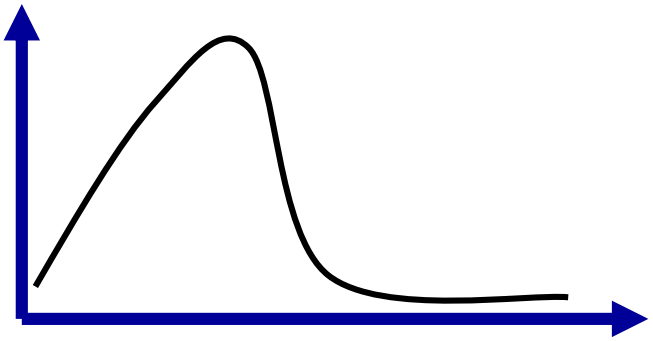
نمودارها-چند ضلعی تجمعی



کجی منفی

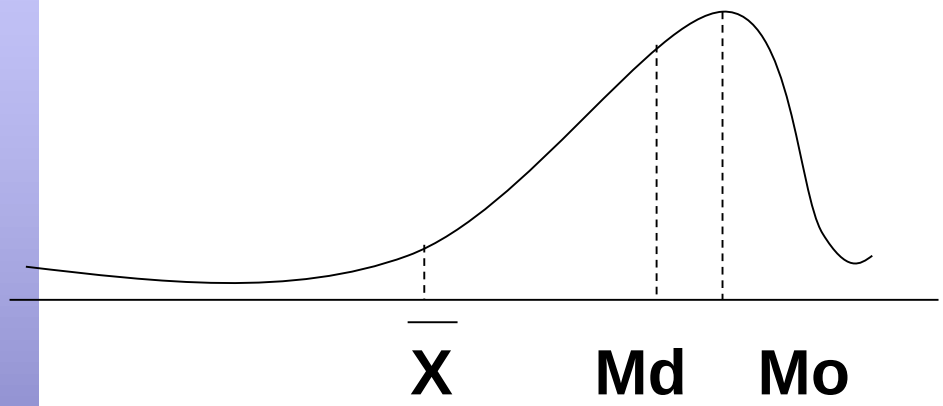


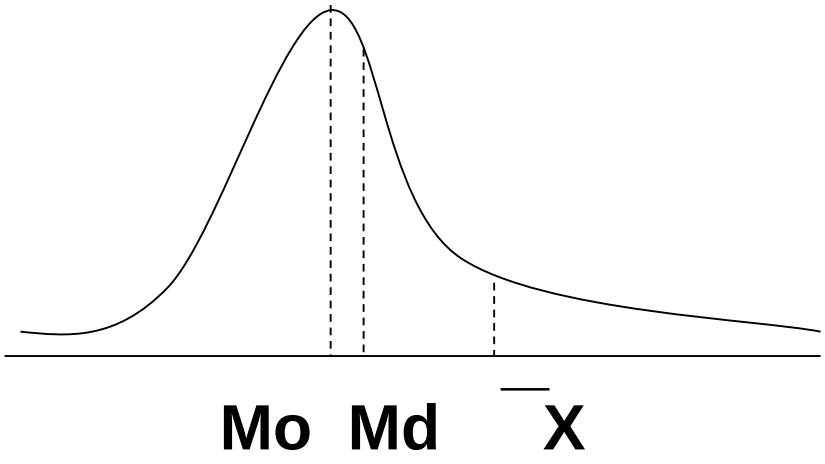
متقارن

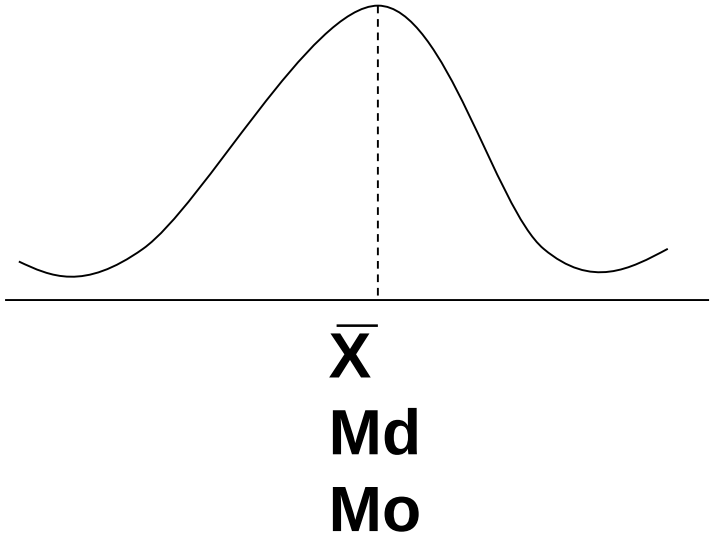


کجی مثبت

مقایسه نما، میانه، میانگین







مقایسه نما، میانه، میانگین

اگر منحنی نرمال باشد

$$MO = Md = \overline{X}$$

اگر کجی مثبت باشد

$$MO < Md < \overline{X}$$

اگر کجی منفی باشد

$$MO > Md > \overline{X}$$

نشان دادن کجی منحنی با استفاده از چارکها

کجی مثبت =

$$Q_3 - Q_2 \rangle Q_2 - Q_1$$

متقارن =

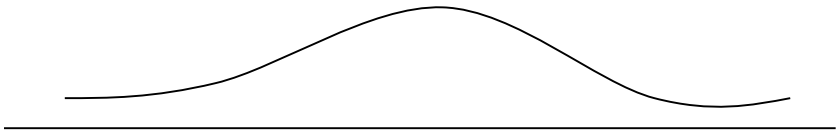
$$Q_3 - Q_2 = Q_2 - Q_1$$

کجی منفی =

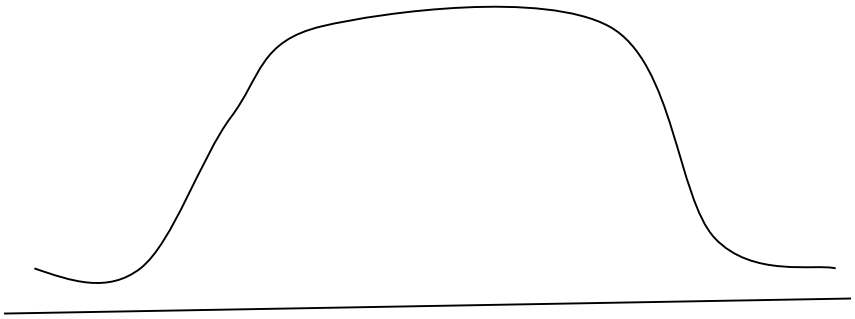
$$Q_3 - Q_2 \langle Q_2 - Q_1$$

کشیدگی

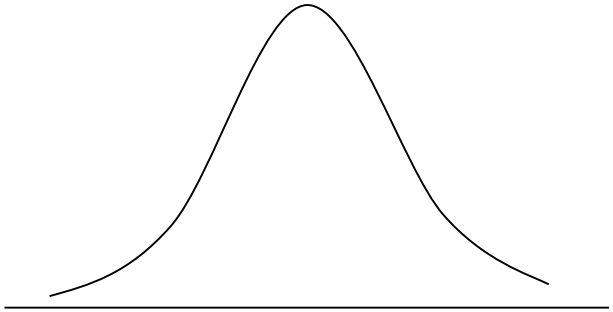
کشیدگی بیانگر نحوه پراکندگی توزیع است



کشیدگی منفی (پراکندگی زیاد)



کشیدگی مثبت(پراکندگی متوسط)



کشیدگی ۰ (پراکندگی خیلی کم)

نمودارها

داده‌های آماری را بوسیله نمودار نشان می‌دهیم. نمودارها به ما کمک می‌کنند که تصویر داده‌ها بهتر دیده شود و رفتار آن‌ها را به راحتی مورد بررسی قرار دهیم. برای رسم نمودارها بسته به مقیاس داده‌ها نمودارهای مختلفی رسم می‌کنیم، چنانچه مقیاس داده‌ها از نوع فاصله‌ای و نسبی باشد از نمودارهای کمی استفاده می‌شود و در صورتی که مقیاس اسمی یا رتبه‌ای باشد از نمودارهای کیفی (وصفی) استفاده می‌کنیم.

نمودارها

- نمودارهای کمی (Numerical Charts)

1. نمودار هیستوگرام (بافت نگار). (Histogram Chart)
2. نمودار چند ضلعی (چند بر فراوانی). (Polygon)
3. نمودار فراوانی تجمعی (اجایو). (Ogive)

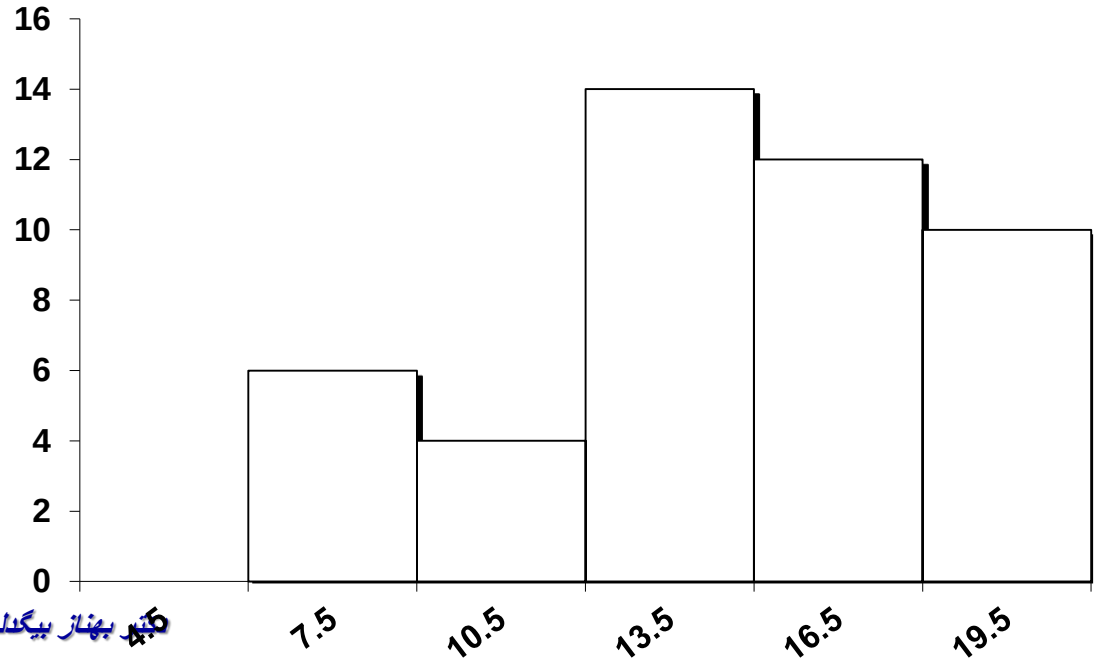
- نمودارهای وصفی

1. نمودار ستونی (میله‌ای). (Bar-chart)
2. نمودار دایره‌ای (کلوچه‌ای). (Pie chart)
3. نمودار پار تو. (Pareto chart)

نمودارهای کمی

۱- نمودار هیستوگرام (بافت نگار).

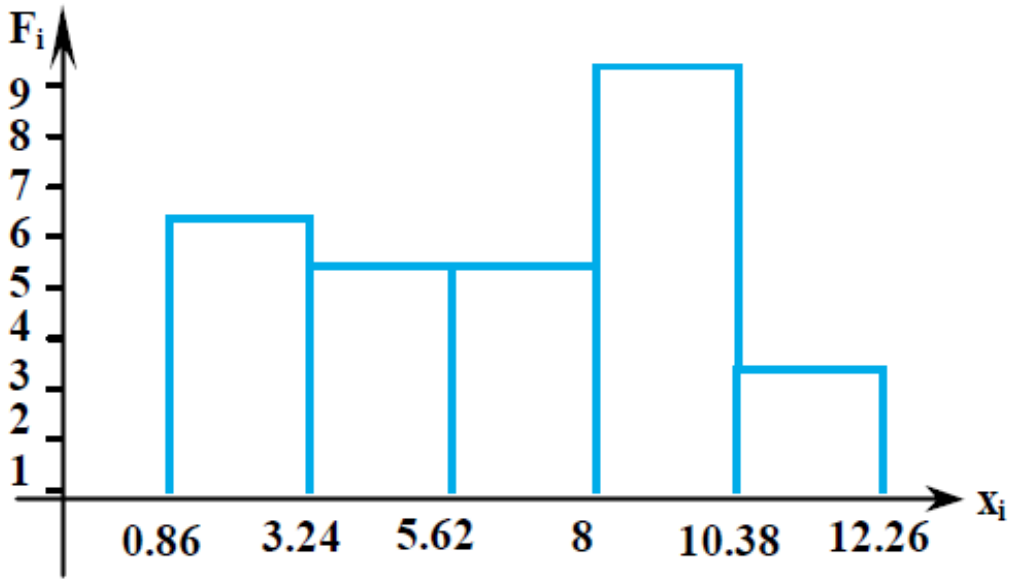
- ✓ برای رسم این نمودار از حدود واقعی طبقات و فراوانی‌های مطلق یا نسبی استفاده می‌شود بر روی محور xها حدود واقعی طبقات و بر روی محور yها فراوانی‌های مطلق یا نسبی قرار می‌گیرد. گاهی اوقات در رسم نمودار هیستوگرام، از مفهوم چگالی نیز استفاده می‌شود. بدین ترتیب که بر روی محور عمودی چگالی مطلق یا نسبی داده‌ها را قرار می‌دهند. چگالی مطلق یا نسبی در هر طبقه از تقسیم فراوانی مطلق یا نسبی آن طبقه بر طول طبقات به دست می‌آید.



نمودارهای کمی

۱- نمودار هیستوگرام (بافت نگار).

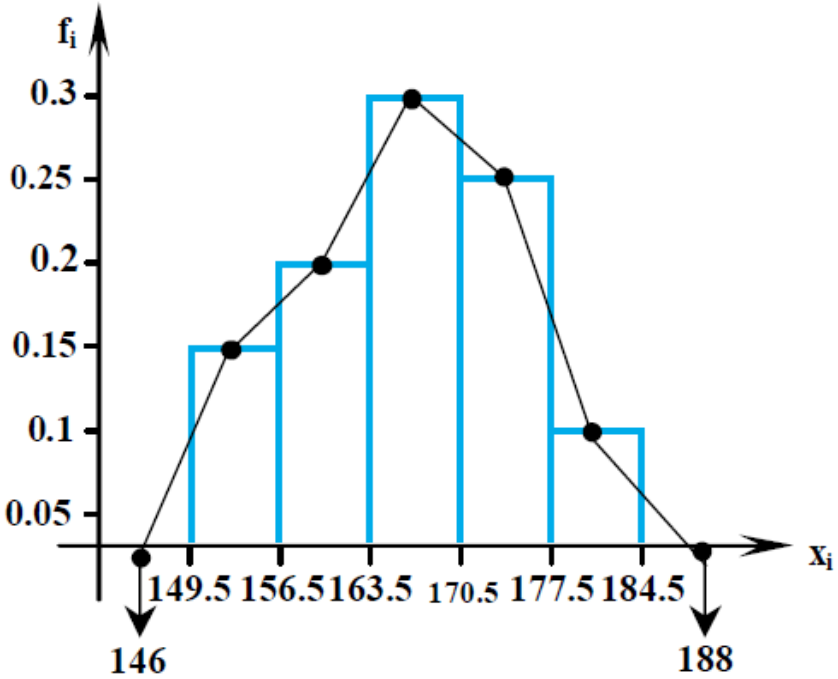
حدود طبقات	۰/۸۶ – ۳/۲۴	۳/۲۴ – ۵/۶۲	۵/۶۲ – ۸	۸ – ۱۰/۳۸	۱۰/۳۸ – ۱۲/۲۶
F_i	۶	۵	۵	۹	۳



نمودارهای کمی

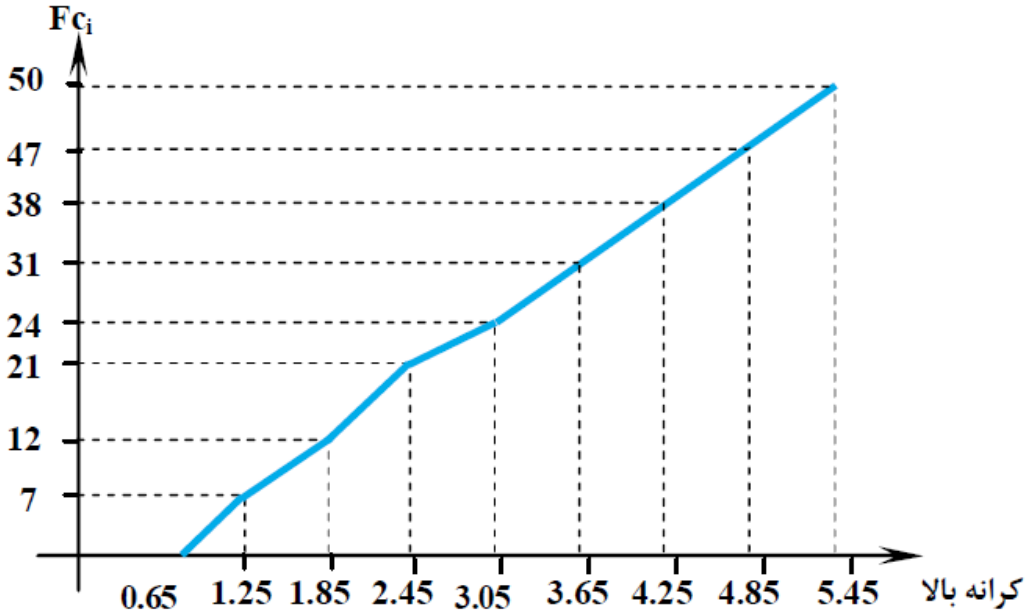
۲- نمودار چند ضلعی (چند بر فراوانی).

اگر وسط مستطیل‌های نمودار هیستوگرام را با پاره خط به یکدیگر متصل کنیم، نمودار چند ضلعی حاصل می‌شود، به عبارت دیگر می‌توان گفت که در رسم این نمودار روی‌ها فراوانی‌های مطلق یا نسبی را قرار می‌دهیم. yها مرکز طبقات و بر روی محور xمحور توجه کنید که برای به دست آوردن ابتدا و انتهای نمودار به اندازه نصف طول دسته از نقطه ابتدایی و انتهایی نزدیک و دور می‌شویم.



۳- نمودار فراوانی تجمعی (اجایو).

به طورکلی دو نوع نمودار فراوانی تجمعی داریم، در نمودار اول محور افقی براساس متوسط طبقات مندرج می‌شود و بر روی محور yها فراوانی‌های تجمعی قرار می‌گیرد که این نمودار به پلی گان فراوانی تجمعی معروف است. در نمودار دوم بر روی محور افقی کرانه‌های بالای هر طبقه و بر روی محورyها فراوانی‌های تجمعی قرار می‌گیرد.



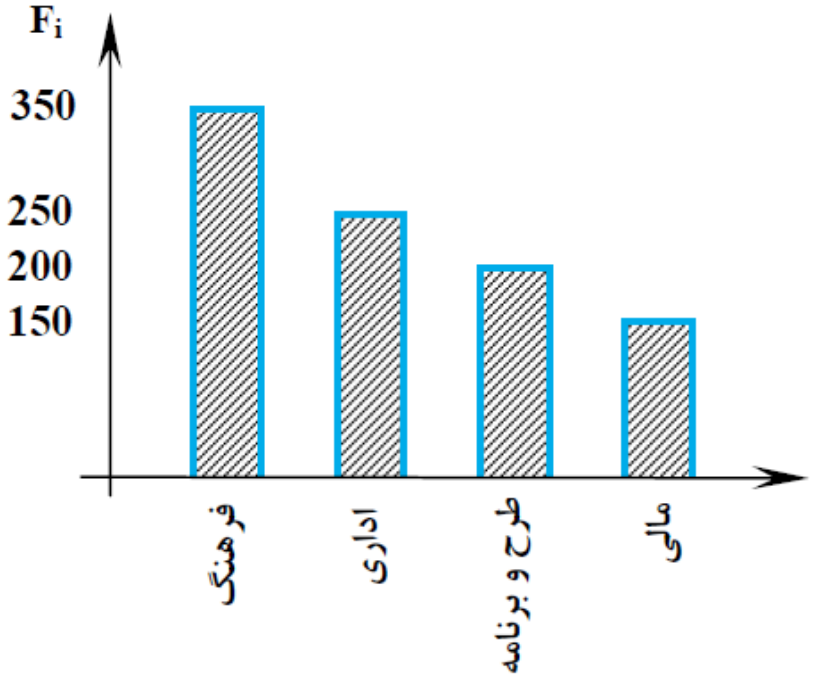
«نمودار فراوانی تجمعی»

نمودارهای وصفی

۱- نمودار ستونی (میله‌ای).

در این نوع نمودار معمولاً محور Xها نشان دهنده کیفیت مشاهدات و محور عمودی آن نشان دهنده فراوانی مطلق یا نسبی هر گروه می‌باشد.

معاونت	اداری	مالی	طرح و برنامه	فرهنگ
تعداد کارکنان	۲۵۰	۱۵۰	۲۰۰	۳۵۰



نمودارهای وصفی

۱- نمودار دایره‌ای (کلوچه‌ای).

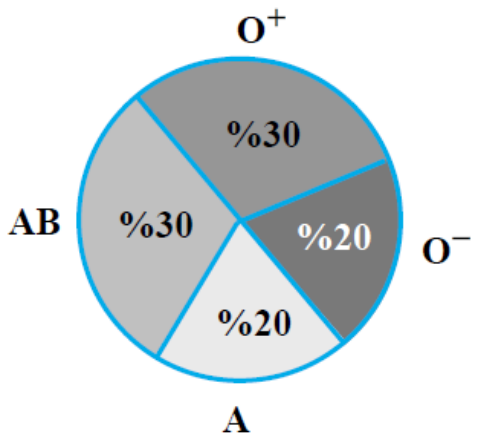
برای رسم این نمودار، ابتدا دایره‌ای رسم کرده سپس قطاع‌هایی برحسب درصد یا درجه از دایره برای هر طبقه جدا می‌کنیم. سهم این قطاع‌ها از دایره به صورت زیر محاسبه می‌گردد:

$$\text{سهم برحسب درصد} = \frac{F_i}{n} \times ۱۰۰ \quad , \quad \text{سهم برحسب درجه} = \frac{F_i}{n} \times ۳۶۰$$

نمودارهای وصفی

۱- نمودار دایره‌ای (کلوچه‌ای).

گروههای خونی	A	AB	O^+	O^-
F_i	۱۰۰۰	۱۵۰۰	۱۵۰۰	۱۰۰۰



$$\text{سهم گروه خونی } A = \frac{F_i}{N} \times ۱۰۰ = \frac{۱۰۰۰}{۵۰۰۰} \times ۱۰۰ = ۲۰\%$$

$$\text{سهم گروه خونی } AB = \frac{F_i}{N} \times ۱۰۰ = \frac{۱۵۰۰}{۵۰۰۰} \times ۱۰۰ = ۳۰\%$$

$$\text{سهم گروه خونی } O^+ = \frac{F_i}{N} \times ۱۰۰ = \frac{۱۵۰۰}{۵۰۰۰} \times ۱۰۰ = ۳۰\%$$

$$\text{سهم گروه خونی } O^- = \frac{F_i}{N} \times ۱۰۰ = \frac{۱۰۰۰}{۵۰۰۰} \times ۱۰۰ = ۲۰\%$$

نمودارهای وصفی

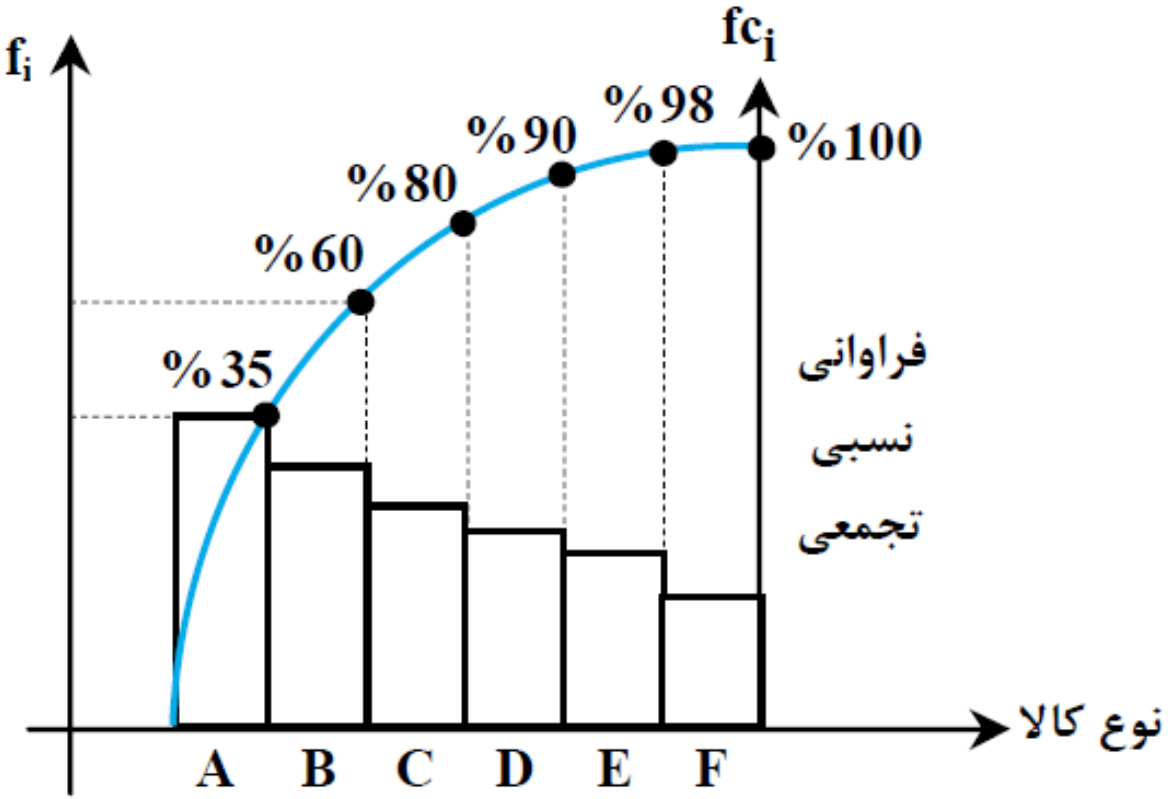
۱- نمودار پار تو.

در نمودار پاره تو، فراوانی نسبی هر موضوع روی محور عمودی و نوع آن روی محور افقی آورده می‌شود. نمودار پارتو همیشه به ترتیب نزولی فراوانی‌ها رسم می‌شود. پر وقوع ترین موضوع در سمت چپ قرار می‌گیرد. توجه کنید که این نمودار سه محور دارد که محور سوم آن نشان دهنده فراوانی‌های نسبی تجمعی می‌باشد. این نمودار برای تحلیل موجودی انبار کالا، نواقص سیستم‌ها، توزیع درآمد و توزیع کارمندان سازمان‌ها کاربرد فراوانی دارد.

نمودارهای وصفی

۱- نمودار پار تو.

کالا	A	B	C	D	E	F
درصد	%۳۵	%۲۵	%۲۰	%۱۰	%۸	%۲



تحلیل اکتشافی داده‌ها

شناسایی انفرادی داده‌ها را تحلیل اکتشافی داده‌ها می‌گویند. نمودارهایی در مراحل اولیه داده‌ها به عنوان نمودارهای تحلیل اکتشافی رسم می‌شوند در اینجا به دو نمونه از آن‌ها اشاره می‌کنیم.

۱- نمودار شاخه و برگ (ساقه و برگ). (Stem and leaf display)

۲- نمودار جعبه‌ای (Box Plot)

تحلیل اکتشافی داده‌ها

۱- نمودار شاخه و برگ (ساقه و برگ).

فرض می‌کنیم $x_1, x_2, \ldots, x_n$ مشاهدات باشند و هر مشاهده دست کم دو رقمی باشد. برای تهیه نمودار شاخه و برگ، ارقام را به دو بخش تقسیم می‌کنیم، شاخه شامل یک یا چند رقم اولیه و برگ آن‌ها شامل رقم‌های باقیمانده می‌باشند.

۰,۱۷,۱۳,۱۳,۱۸,۱۰,۱۹,۵,۷,۴,۱,۲۱,۲۵,۲۴,۲۹,۳۵

شاخه	برگ
۰	۵,۴,۷,۰,۱
۱	۷,۳,۳,۸,۹
۲	۱,۵,۴,۹
۳	۵

تحلیل اکتشافی داده‌ها

۲- نمودار جعبه‌ای

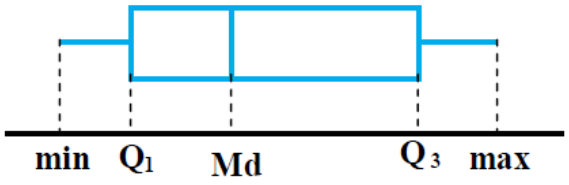
مراحل رسم این نمودار به شرح زیر است:

1) داده‌های خام را به صورت صعودی مرتب می‌کنیم سپس یک خط افقی را چنان مدرج می‌کنیم که بتوان همه داده‌ها را روی آن نشان داد.

2) میانه و چارک‌های اول و سوم داده‌ها را محاسبه می کنیم.

3) بالای خط مدرج شده مستطیلی رسم می‌کنیم که طول آن برابر با Q_3-Q_1 بوده و از نقطه Q_1 شروع و به نقطه Q_3 ختم شود. عرض مستطیل را بصورت معقول در نظر گرفته شود. این مستطیل را جعبه یا باکس می‌نماییم.

4) اندازه میانه را به صورت خطی به موازات عرض مستطیل رسم نموده و مستطیل را به وسیله یک خط منقطع به موازات خط مدرج شده به دو قسمت تقسیم می‌کنیم.



۲- نمودار جعبه‌ای

5) مرزهای داخلی و خارجی داده‌ها به صورت زیر تعیین می شوند:

مرز داخلی پایین $= Q_۱ - ۱/۵(Q_۳ - Q_۱)$

مرز داخلی بالا $= Q_۳ + ۱/۵(Q_۳ - Q_۱)$

مرز خارجی پایین $= Q_۱ - ۳(Q_۳ - Q_۱)$

مرز خارجی بالا $= Q_۳ + ۳(Q_۳ - Q_۱)$

6) هر داده‌ای که خارج از مرزهای داخلی قرار گرفته باشد را یک داده پرت می‌نامند و چنانچه بین مرزهای داخلی و خارجی قرار گیرد آن را داده پرت ضعیف نامیده و با علامت O نشان می‌دهیم و چنانچه خارج از مرزهای خارجی قرار گیرد آن را داده پرت قوی نامیده و با علامت • نشان می‌دهیم.

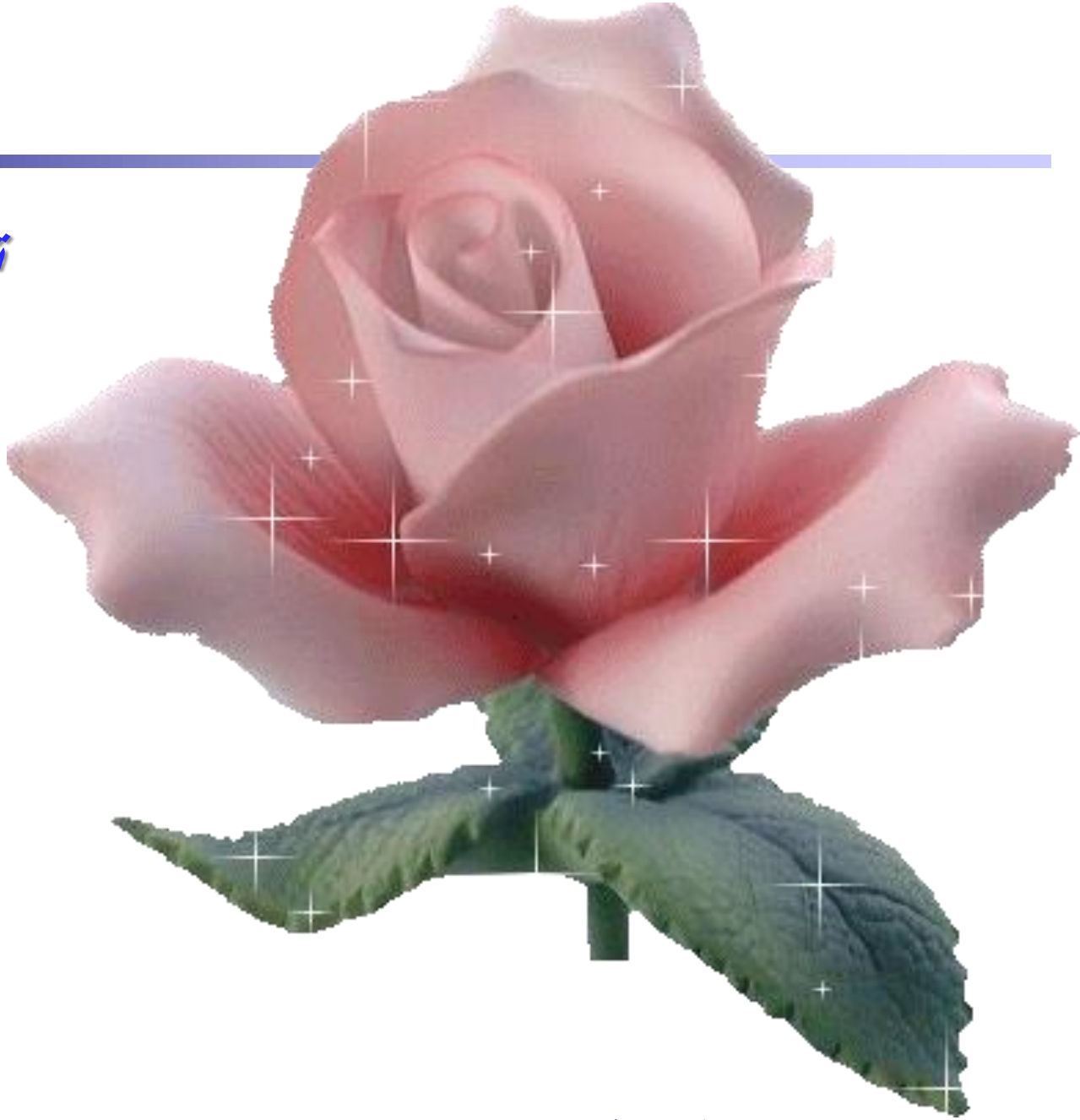
تحلیل اکتشافی داده‌ها

۲- نمودار جعبه‌ای

با استفاده از نمودار جعبه‌ای می‌توان اطلاعات زیر را در مورد داده‌ها کسب نمود:

۱- اگر میانه نزدیک وسط مستطیل (جعبه) باشد توزیع داده‌ها تقریباً متقارن است.

۲- اگر میانه در طرف چپ وسط مستطیل باشد توزیع چولگی به راست و اگر میانه در طرف راست وسط مستطیل قرار گیرد، توزیع چولگی به چپ است.

۳- در مقایسه نمودار جعبه‌ای دو مجموعه از داده‌ها می‌توان پراکندگی آن‌ها را با توجه به طول مستطیل‌های نمودار با یکدیگر مقایسه نمود. مستطیلی که طول بزرگتری دارد، داده‌های آن دارای پراکندگی بیشتر می‌باشد.

تشکر از توجه شما