آمار و احتمالات مهندسی

آمار توصیفی

دکتر بهناز بیگدلی

دانشکده مهندسی عمران

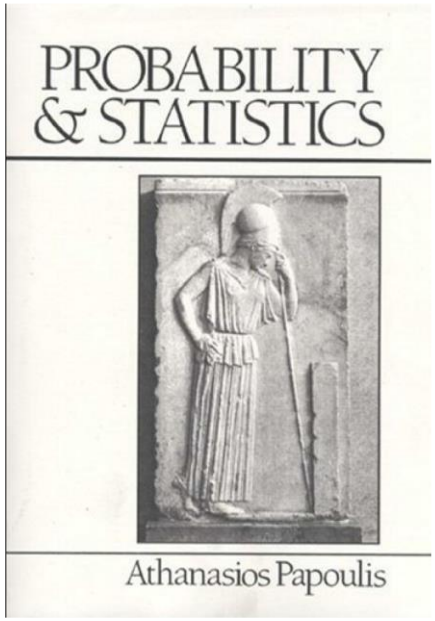
دانشگاه صنعتی شاهرود

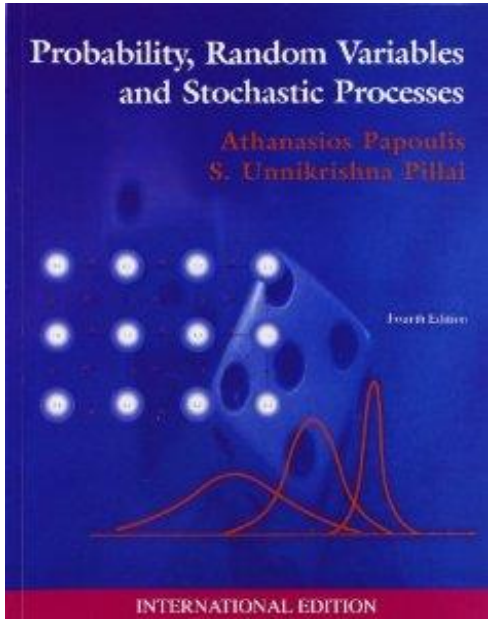
ساختار درس

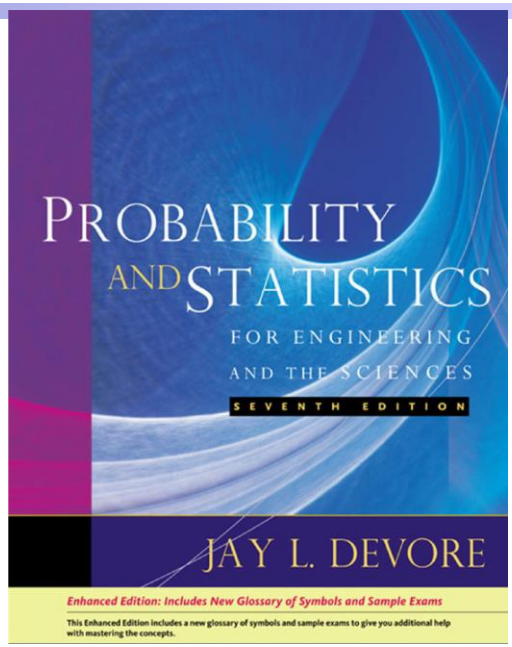
نحوه ارزیابی

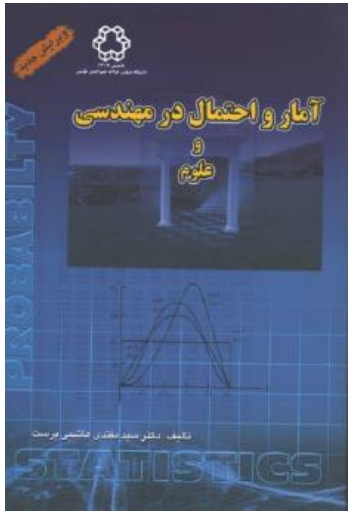
- ✓ امتحان پایان ترم (۱۴)
- ✓ امتحان کلاسی (۲)
- ✓ تمرینات (۴)

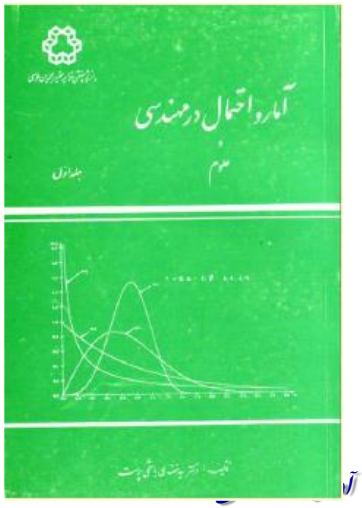
منابع


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

آمار توصیفی

دراین فصل مسائل زیر بررسی می شود:

- ✓تعریف علم آمار
- ✓-مفاهیم اساسی
- ✓-شاخص های گرایش مرکزی
- ✓-شاخص های پراکندگی
- ✓-جدول توزیع فراوانی
- ✓-نمودارها
- ✓-چولگی و برجستگی
- ✓-جامعه آماری دو بعدی

تعریف علم آمار

علمی که به کمک آن می توان آگاهی و اطلاعات مربوط به یک موضوع را مورد تجزیه و تحلیل قرار داد، «علم آمار» نامید می‌شود.

روش‌های آماری دارای دو وظیفه مهم هستند:

۱- به پژوهشگر در طبقه‌بندی، خلاصه کردن، توصیف و تفسیر و برقراری ارتباط از طریق اطلاعات جمع‌آوری شده کمک می کند.

۲- به پژوهشگران امکان می‌دهد که با استفاده از اطلاعات جمع‌آوری شده از نمونه کوچکی از آزمودنی‌ها ویژگی‌های جامعه‌ای را که نمونه از آن انتخاب شده است برآورد یا استنباط کنند.

انواع روش‌های آماری
مورد استفاده در وظایف اول و دوم

۱- آمار توصیفی: مجموعه روش‌هایی است که به خلاصه کردن، طبقه‌بندی، توصیف و تفسیر داده‌ها می‌پردازد.

۲- آماراستنباطی: مجموعه روش‌هایی است که معمولاً برای بیان رابطه بین دو یا چند متغیر و تعمیم ویژگی‌های نمونه آماری به جامعه آماری به کار برده می‌شوند.

نکات

- ✓ اولین و مفیدترین قدم در سازمان دادن به دادها ، مرتب کردن آنها بر اساس یک ملاک منطقی است.
- ✓ با استفاده از روش‌های آمار توصیفی می‌توان دقیقاً ویژگی‌های یک دسته از اطلاعات را بیان کرد.
- ✓ روش‌های آمار توصیفی همیشه برای تعیین و بیان ویژگی‌های اطلاعاتی به کار برده می‌شود که به وسیله پژوهشگران جمع آوری شده‌اند.
- ✓ هدف نهایی آمار استنباطی برآورد ویژگی‌های جامعه است.

دلایل مطالعه آمار

۱- کاربرد روزانه:

در تسریع برنامه‌ها، تهیه آزمون و تفسیر کمک می کند. (مثلا : تفسیر نمره‌ها توسط معلم، تفسیر مشاهدات توسط روانشناسان، ارزشیابی اطلاعات و تعمیم آنها توسط جامعه شناسان و ...)

۲- حل مسائل:

پژوهش غالباً در یک مقیاس محدود و به منظور کشف اطلاعات ضروری برای حل مسائل عملی انجام می‌شود. در زندگی موضوعات و مسائل مختلفی وجود دارد که از طریق آمار به آنها پاسخ داده می‌شود.

۳- پژوهش نظریه‌ایی:

از طریق آمار می توان نظریه ها(مثلا: روانشناسی، تربیتی و جامعه شناسی) را مورد آزمون قرار داد.

۴- کاربرد پژوهش و درک و فهم آن:

برای درک نتایج پژوهشی و تشخیص اینکه روش‌های آماری درست انتخاب و تفسیر شده‌اند خواننده باید با روش‌های آماری آشنایی داشته باشد.

مفاهیم اساسی

۱-جامعه

۲-نمونه

۳-آماره، پارامتر

۴-انواع متغیرها

۱-جامعه

هدف نهایی آمار استنباطی برآورد ویژگی‌های جامعه است. جامعه فقط به گروهی از افراد محدود نمی‌شود. ممکن است شامل تمام روش‌های آزمایشگاهی، انواع نان، انواع محصولات صنعتی و غیره باشد.

به گروهی از افراد، اشیاء، حوادث و ... که حداقل دارای یک صفت مشترک باشند جامعه آماری گفته می‌شود. در یک تحقیق کلیه افراد یا واحدهای مورد بررسی جامعه آماری می‌باشند.

مثلاً در نقشه‌برداری، از یک منطقه برای تعیین مختصات مکان‌های مختلف کلیه قرائت‌ها شامل، طول، زاویه، آزیموت و دیگر مشاهدات جامعه آماری می‌باشند.

$$X_1, X_2, \ldots, X_i, \ldots, X_N$$

که x_i عضو iام جامعه است برای $i=1,2,\ldots,N$

۲- نمونه

اگر اطلاعات مربوط به همه افراد جامعه جمع آوری شود، «روش سرشماری» صورت گرفته، ولی اگر اطلاعات برای برخی از افراد جامعه جمع آوری شود «روش نمونه برداری» انجام شده است.

نمونه عبارتست از زیر جامعه‌ایی که از کل جامعه انتخاب می‌شود و معرف آنست. دلیل انتخاب نمونه اینست که اندازه‌گیری ویژگی مورد پژوهش برای تک تک افراد یا عناصر جامعه غیر ممکن است.

نمونه باید نماینده واقعی جامعه باشد. مراد از نماینده واقعی بودن اینست که بین ویژگی‌های نمونه و جامعه شباهت تقریباً کاملی وجود داشته باشد. معرف یا نماینده واقعی بودن نمونه غالباً از طریق انتخاب تصادفی نمونه امکانپذیر است.

۳- آماره و پارامتر

اندازه‌هایی که از نمونه به دست می‌آیند آمار یا آماره نامیده می‌شوند و اندازه‌هایی که ویژگی‌های جامعه را تعیین می‌کند پارامتر نامیده می‌شوند. معمولا آماره با حروف انگلیسی و پارامتر با حروف یونانی نوشته می‌شود.

ویژگی شاخص	میانگین	واریانس	انحراف استاندارد	نسبت	همبستگی	تعداد مشاهدات
آمار	$\bar{X}$	S^2	s	P	r	n
پارامتر	μ	σ^2	σ	-	-	N

۴-متغیر

الف: انواع متغیر از نظر ماهیت

ب: انواع متغیر از نظر نقشی که در پژوهش دارد

ج: انواع متغیر از نظر فواصل بین اعداد

۴-متغیر

الف: انواع متغیر از نظر ماهیت

۱- متغیر کمی: به متغیرهایی اطلاق می‌شوند که از نظر مقدار یا ارزش متفاوت هستند و به صورت عدد نوشته می شوند مانند سن، نمرات درسی و ...

۲- متغیر کیفی: هر متغیری که نتوان آن را بصورت عددی نمایش داد مانند :جنس، رنگ مو، مذهب.

۴-متغیر

ب: انواع متغیر از نظر نقشی که در پژوهش دارد

۱- متغیر مستقل: متغیری است که بر متغیرهای دیگر اثر می‌گذارد، متغیر پیش فرض است و از طریق آن متغیر وابسته اندازه‌گیری و تعیین می‌شود. در تحقیق آزمایشی متغیر مستقل متغیری است که توسط محقق دستکاری می‌شود تا تأثیرش بر متغیر وابسته مشخص شود.

۲- متغیر وابسته: متغیری است که ارزش یا مقدار آن به متغیر مستقل بستگی دارد. متغیر وابسته در اختیار محقق نیست و محقق نمی‌تواند در آن دخل و تصرف و دستکاری به عمل آورد.

۴-متغیر

ج: انواع متغیر از نظر فواصل بین اعداد

اگر متغیر را در مورد تک تک افراد جمعیت یا نمونه‌ای از آن با مقیاسی مناسب اندازه‌گیری کنیم یک مجموعه از اعداد به دست می‌آید که آن را داده می‌نامند . داده‌ها دو نوع هستند:

۱- داده‌های گسسته : متغیری که فاصله بین اعداد را درنظر نمی‌گیرد و ارزش‌های موجود بین دو مقدار دارای معنی نیست. مانند : تعداد دانشجویان، تعداد معلمان...

۲- داده‌های پیوسته: متغیری که هر ارزش یا مقداری (کسری، اعشاری) را می‌توان به آن اختصاص داد مانند: قد، وزن و...

۴-متغیر

ج: انواع متغیر از نظر فواصل بین اعداد

در عمل تشخیص بین متغیر پیوسته و گسسته به صورت نظری امکان پذیر نیست. دلیل این امر فقدان وسایل اندازه گیری دقیق و مناسب است. در پژوهش غالباً متغیرهایی که ذاتاً پیوسته هستند به صورت گسسته مورد بحث قرار می‌گیرند.

مثلاً سن (پیوسته) به دلیل طبقه بندی کردن افراد به متغیر گسسته تبدیل می‌شود.

شاخص‌های گرایش مرکزی:

۱-میانگین

۲- میانه

۳- نما

۴- چندک‌ها

۱- میانگین

الف-میانگین حسابی

ب- میانگین هندسی

پ-میانگین هارمونیک

ت- میانگین پیراسته

الف- میانگین حسابی

فرض کنید جامعه مورد بررسی دارای N عضو $X_1, X_2, \ldots, X_n$ باشد. میانگین جامعه از رابطه زیر بدست می آید.

$$\mu = \overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i = \frac{1}{N}(X_1 + X_2 + \ldots + X_N)$$

الف-میانگین حسابی در جدول توزیع فراوانی

X	f	fx
18	1	18
17	2	34
15	2	30
12	3	36
11	5	55
	13	173

$$\overline{X} = \frac{\Sigma fX}{N}$$

$$\overline{X} = \frac{173}{13} = 13 \quad / \ 3$$

الف-میانگین حسابی در اعداد طبقه بندی شده

X	f	X'	fX'
5 -9	3	7	21
10 -14	7	12	84
15 -19	5	17	85
20-24	5	22	110
	$\Sigma = 20$		$\Sigma = 300$

$$\overline{X} = \frac{\Sigma fX'}{N}$$

$$\overline{X} = \frac{300}{20} = 15$$

ب- میانگین هندسی

اگر $X_1, X_2, \ldots, X_n$ یک نمونه به حجم n از جامعه مورد بررسی باشد میانگین هندسی از رابطه زیر بدست می آید و با علامت G نمایش داده می شود.

$$G = \sqrt[n]{x_1, x_2, \ldots, x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i}$$

$$\log G = \frac{1}{n}\sum_{i=1}^{n} \log X_i \Longrightarrow G = antilog\left(\frac{1}{n}\sum_{i=1}^{n} \log X_i\right)$$

ب- میانگین هندسی با فراوانی

اگر $X_1, X_2, \ldots, X_n$ یک نمونه به حجم n از با فراوانی‌های $f_1, f_2, \ldots, f_n$ از جامعه مورد بررسی باشد میانگین هندسی از رابطه زیر بدست می‌آید و با علامت G نمایش داده می‌شود.

$$G = \sqrt[n]{x_1^{f_1}, x_2^{f_2}, \ldots, x_n^{f_n}} = \sqrt[n]{\prod_{i=1}^{n} x_i^{f_i}}$$

$$\log G = \frac{1}{N}\sum_{i=1}^{n} f_i \log X_i \Longrightarrow G = antilog\left(\frac{1}{N}\sum_{i=1}^{n} f_i \log X_i\right)$$

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

پ-میانگین هارمونیک

اگر $X_1, X_2, \ldots, X_n$ یک نمونه به حجم n از جامعه مورد بررسی باشد میانگین هارمونیک از رابطه زیر بدست می آید و با علامت H نمـایش داده می شود.

$$H = \frac{1}{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}} = \frac{1}{\frac{1}{N}\sum_{i=1}^{n}\frac{f_i}{x_i}}$$

$$\frac{1}{H} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i} = \frac{1}{N}\sum_{i=1}^{n}\frac{f_i}{x_i}$$

ت-میانگین پیراسته

اگر k تا از مشاهدات حذف شده باشند میانگین پیراسته از رابطه زیر بدست می آید . k<n

$$\bar{x}_p = \frac{1}{n-k}\sum_{i=1}^{n-k} x_i$$

رابطه بین میانگین ها

- ✓ همیشه بین میانگین‌های
- ✓ حسابی μ
- ✓ هندسی G
- ✓ هارمونیک H
- ✓ رابطه زیر برقرار است.

$$\mu > G > H$$

ویژگی‌های میانگین

۱- به تک تک اعداد توزیع فراوانی حساس است.

۲- دارای ثبات بیشتری نسبت به میانه و نما است.

۳- میانگین در نمونه های مختلف یک جامعه بیشتر از نما و میانه به همدیگر نزدیک می‌باشند.

۴-اگر تمام داده‌ها با عدد ثابتی جمع یا تفریق یا ضرب یا تقسیم شوند میانگین در آن عدد جمع، تفریق، ضرب و تقسیم می‌شود.

۵- مجموع انحراف داده‌ها از میانگین برابر صفر است.

$$\Sigma(X-\overline{X})=0$$

۶- مجموع مجذورهای انحراف داده‌ها از میانگین همیشه کوچکتر یا مساوی با مجموع مجذور انحراف نمره ها از هر عدد دیگری است.

$$\Sigma(X-\overline{X})^2 \le \Sigma(X-A)^2$$

میانگین‌های وزن‌دار

✓ اگر مقدار x_i در یک سری از داده‌ها دارای وزن w_i باشد در این صورت خواهیم داشت.

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{n} w_i x_i \qquad N = \sum_{i=1}^{n} w_i$$

$$\log G = \frac{1}{N}\sum_{i=1}^{n} w_i \log X_i \;==\gg G = antilog\left(\frac{1}{N}\sum_{i=1}^{n} w_i \log X_i\right)$$

$$H = \frac{1}{\frac{1}{N}\sum_{i=1}^{n}\frac{w_i}{x_i}}$$

۲- میانه

میانه عددی است که ۵۰٪ داده‌ها پایین تر و ۵۰٪ داده‌ها بالاتر از آن قرار دارند.

✓ ویژگی ها:

الف- میانه مشاهدات را به دو بخش مساوی تقسیم می کند.

ب- منحصر به فرد است.

ج- تحت تأثیر داده‌های پرت قرار نمی‌گیرد.

د- محاسبه آن ساده است.

۱- میانه در سری داده‌ها:

$$\mathrm{Md} = X_{\frac{n+1}{2}}$$

در سری داده‌ها میانه داده ی شماره $\frac{n+1}{2}$ است:

۲- میانه در توزیع ساده

اگر n فرد باشد

$$\mathrm{Md} = X_{\frac{n+1}{2}}$$

اگر n زوج باشد

$$\mathrm{Md} = \frac{X_{\frac{n}{2}+1} + X_{\frac{n}{2}}}{2}$$

۳- نما

-نما یا مد داده‌ای است که فراوانی تکرار آن از سایر داده‌ها بیشتر است.

-نما را با علامت Mo نشان می‌دهند.

-در سری داده‌ها نما داده‌ای است که بیش از سایر داده‌ها تکرار شده است.

در یک توزیع گروه‌بندی شده نما در گروهی جای دارد که دارای بیشترین فراوانی است.

d_1 : اختلاف فراوانی گروه شامل نما با گروه ماقبل خود

$d_۱ = f_i - f_{i-۱}$, $d_۲ = f_i - f_{i+۱}$

d_2 : اختلاف فراوانی گروه شامل نما با گروه مابعد خود

$$Mo = a_i + \frac{d_1}{d_1 + d_2} \times c$$

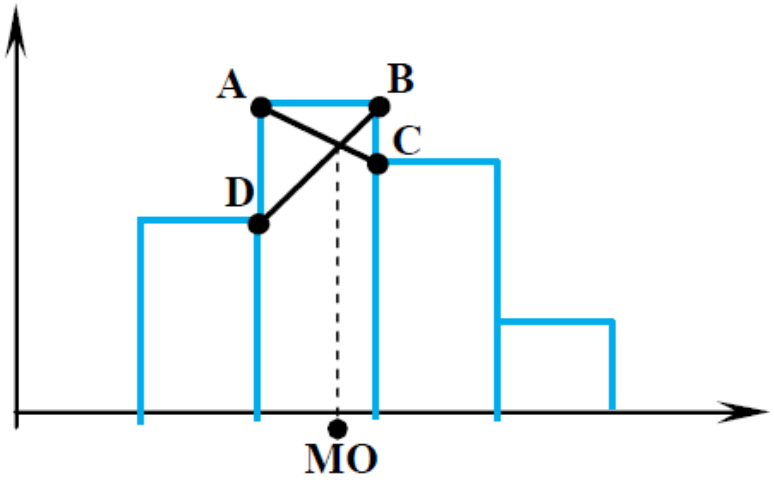
C:طول کلاس طبقه نمادار

ai: حدپایین کلاس نمادار

۳- نما

- ***خواص نما یا مد داده‌ها***

1. در بین شاخص‌های مرکزی مد اهمیت زیادی ندارد ولی در داده‌های کیفی، مد تنها شاخص مرکزی است. به طور مثال در نظرسنجی های و رأی گیری‌ها که داده‌ها به صورت خوب، بد، متوسط (کیفی) می باشند از مد استفاده می‌شود.
2. مد از نظر هندسی: ابتدا نمودار هیستوگرام را رسم می‌کنیم مطابق شکل بلندترین مستطیل را در نظر گرفته نقطه A را به C و نقطه B را به D متصل می‌کنیم از محل تلاقی این دو خط، خطی عمود بر محورx ها رسم می‌کنیم پای عمود مکان هندسی مد است.
3. هرگاه تک تک داده‌ها را در عددی ثابت ضرب یا بر آن تقسیم کنیم و با عدد ثابت دیگری جمع یا تفریق کنیم همین تغییرات را برای مد نیز خواهیم داشت.



$$MO(ax_i \pm b) = aMO(x_i) \pm b$$

۳- نما

مد/مدال یک توزیع مقداری است از متغییر است که دارای فراوانی ماکزیمم باشد.

در مورد توزیع فراوانی‌های دسته بندی شده، مد با استفاده از رابطه زیر بدست می‌آید.

$$Mod = a_i + \frac{f_i - f_{i-1}}{2f_i - f_{i-1} - f_{i+1}} c$$

a_i حد پایین دسته‌ای که مد در آن قرار دارد.

f_i فراوانی دسته مد

f_{i-1} فراوانی دسته ماقبل دسته مد

f_{i+1} فراوانی دسته مابعد دسته مد

c فاصله دسته

۴-چندک‌ها

مقادیری می‌باشند که دامنه تغییرات داده‌هایی را که به صورت غیر نزولی مرتب شده‌اند به فواصل مورد نیاز تقسیم می‌کنند به طوری که فراوانی‌ها در هر یک از فواصل، درصد خالصی از کل فراوانی را تشکیل می‌دهند.

الف: چارک‌ها

ب: دهک‌ها

ج:صدک‌ها

۴-چندک‌ها

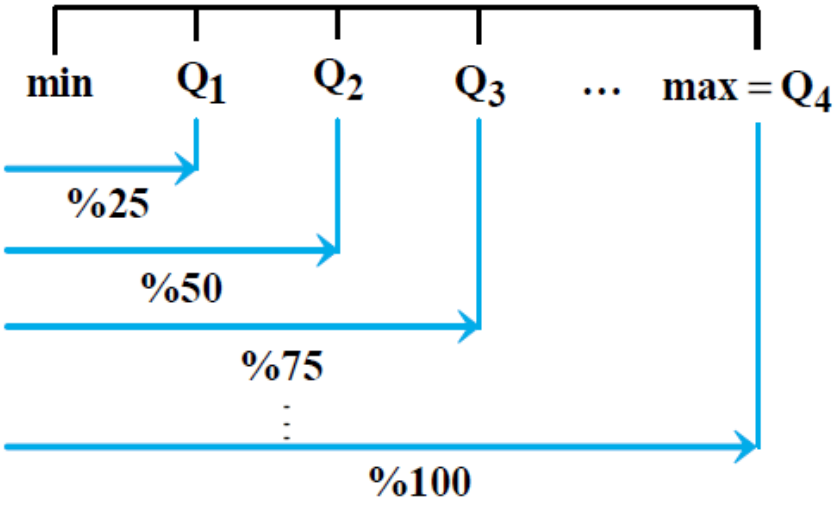
الف: چارک‌ها:

اگر دامنه داده ها را به چهار قسمت مساوی تقسیم کنیم:

چارک اول که با Q_1 نمایش داده می‌شود مقداری است که ٪۲۵ مشاهدات کمتر یا مساوی از آن و ٪۷۵ بالاتر از آن قرار می‌گیرند.

چارک دوم که آن را با Q_2 نشان می‌دهند مقداری است که ٪۵۰ درصد داده ها کمتر یا مساوی از آن و ٪۵۰ بالاتر از آن قرار می‌گیرند و

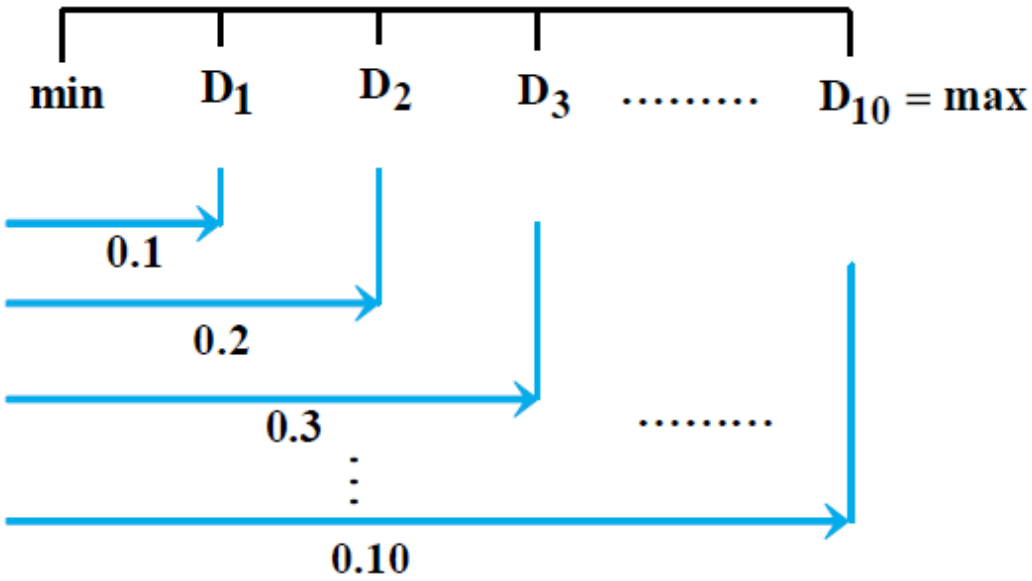
چارک سوم که با Q_3 نشان داده می‌شود مقداری است که ٪۷۵ مشاهدات کمتر یا مساوی از آن و ٪۲۵ بالاتر از آن قرار می‌گیرند.



۴-چندک‌ها

ب: دهک‌ها:

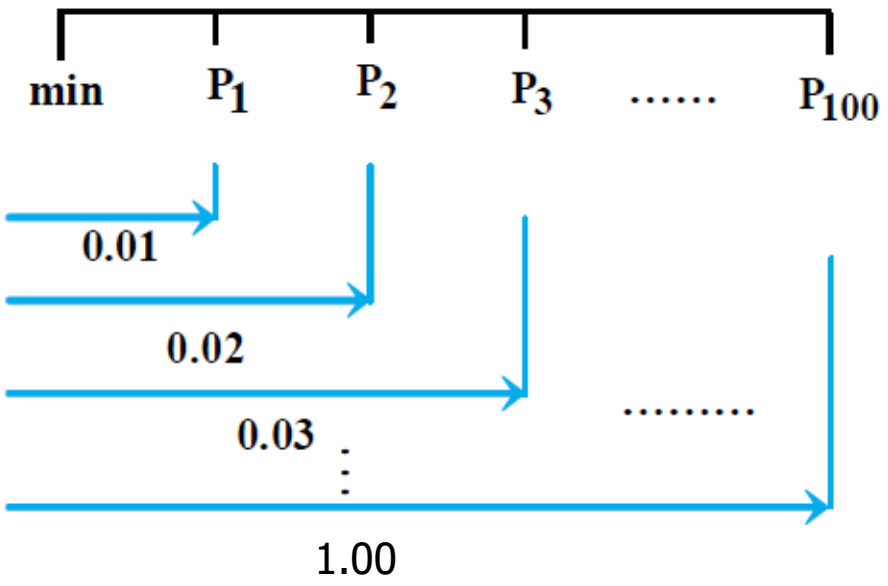
اگر دامنه داده‌ها رابه ده قسمت مساوی تقسیم کنیم، مانند چارک‌ها مقادیر دهک‌ها مشخص می‌شود به طور مثال دهک اول که آن را با D_1 نشان می‌دهند مقداری است که ۱۰٪ داده‌ها پایین تر یا مساوی از آن و ۹۰٪ بالاتر از آن قرار می‌گیرند.



۴-چندک‌ها

ج: صدک‌ها:

اگر دامنه داده‌ها را به صد قسمت مساوی تقسیم کنیم مانند چارک‌ها و دهک‌ها مقادیر صدک‌ها مشخص می‌شود. به طور مثال صدک اول که آن را با P_1 نشان می‌دهند مقداری است که ۱٪ داده‌ها پایین تر یا مساوی از آن و ۹۹٪ بالاتر از آن قرار می‌گیرند.



۴-چندک‌ها

توجه

چارک دوم همان دهک پنجم، همان صدک پنجاهم و همان میانه است.

$$Q_2 = D_5 = P_{50} = Md$$

۴-چندک‌ها

محاسبه چندک‌ها در داده‌های گسسته

فرض کنید n داده را بصورت غیر نزولی مرتب کرده‌ایم برای محاسبه صدک p ۱۰۰ام، ابتدا مقدار (N+1).p را محاسبه می‌کنیم اگر این مقدار عدد صحیح مانند k شود در این صورت x_k صدک p۱۰۰ام ولی اگر مقدار (N+1).p عددی اعشاری شود مقدار صحیح آن را k و اعشار آن را r می‌نامیم و از رابطه زیر صدک مورد نظر را به دست می‌آوریم.

$$(۱-r).x_k + r.x_{k+۱} = x_k + r(x_{k+۱} - x_k)$$

چارکها

۱) داده‌ها را به طور صعودی مرتب می‌کنیم.

۲) داده‌های مرتب شده را از ۱ تا n کدگذاری می‌کنیم.

۳) جایگاه چارک a ام را توسط رابطه‌ی زیر به دست می‌آوریم:

$$CQ_a = \frac{n.a}{۴} + \frac{۱}{۲} \qquad a = ۱, ۲, ۳$$

۴) با استفاده از جایگاه چارک، مقدار چارک را تعیین می‌کنیم.

۴-چندک‌ها

مثال: *در داده‌های زیر دهک هفتم کدام است؟*

5,7,3,4,4,8,9,9, 10

حل:

ابتدا داده‌ها را به صورت غیرنزولی مرتب می‌کنیم

۳,۴,۴,۵,۷,۸,۹,۹,۱۰

$$\begin{cases} N = ۹ \\ p = ۰/۷ \end{cases} \Rightarrow (N+۱)\cdot p = (۹+۱)\times \frac{۷}{۱۰} = ۷ \Rightarrow D_۷ = x_۷ = ۹$$

۴-چندک‌ها

محاسبه چندک‌ها در داده‌هایی با مرکز دسته

در صورتی که مشاهدات آماری گسسته به صورت جدول فراوانی داده شده باشند برای محاسبه چندک‌ها مراحل زیر را به ترتیب اجراء می‌کنیم:

۱) مرکز دسته‌ها (داده‌ها) را به صورت غیر نزولی مرتب می‌کنیم.

۲) مقدار (N+1).p را مشخص می‌کنیم.

۳) فراوانی‌های تجمعی (نسبی) را تشکیل می‌دهیم.

۴) اولین دسته‌ای که فراوانی تجمعی (نسبی) آن بزرگتر یا مساوی (N+1).p باشد محل چندک است.

مثال: در جدول زیر چارک سوم کدام است؟

x_i	۴	۲	۵	۷
F_i	۷	۳	۶	۴

حل:

x_i	۲	۴	۵	۷
F_i	۳	۷	۶	۴
FC_i	۳	۱۰	۱۶	۲۰

$$N = \sum F_i = ۲۰$$

$$(N+۱).P = (۲۰+۱) \times \frac{۳}{۴} = ۱۵/۷۵$$

پس جواب x=5 است.

۴-چندک‌ها

محاسبه چندک‌ها در داده‌های پیوسته

برای محاسبه چندک‌ها در داده‌های پیوسته مراحل زیر را به ترتیب اجراء می‌کنیم.

۱) ابتدا مقدار N.p را مشخص می‌کنیم.

۲) اولین طبقه‌ای که فراوانی تجمعی آن بزرگتر یا مساوی N.p باشد طبقه چندک موردنظر است.

۳) سپس از رابطه زیر مقدار آن را محاسبه می‌کنیم:

$$\text{چندک موردنظر} = L + \left(\frac{N.p - FC_{i-1}}{F_i}\right).I$$

L: کران پایین واقعی طبقه موردنظر

FC_{i-1}: فراوانی تجمعی طبقه قبل

F_i: فراوانی مطلق طبقه موردنظر

I: طول طبقه

رابطه تجربی

✓ در توزیع های تقریبا متقارن، رابطه زیر برقرار است.

$$\mu - Mo = 3(\mu - Md)$$

(میانه – میانگین)۳ = مد – میانگین

جدول توزیع فراوانی

طول کلاس :

$$l_i = b_i - a_i \qquad i = 1,2,\ldots,k$$

محاسبه میانگین و واریانس در جدول توزیع فراوانی :

میانگین حسابی

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

میانگین هندسی

$$G = \sqrt[n]{x_1^{f_1} . x_2^{f_2} \ldots\ldots x_k^{f_k}}$$

میانگین هارمونیک

$$H = \frac{n}{\sum_{i=1}^{k} \frac{f_i}{x_i}}$$

واریانس

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

محاسبه پارامترهای جدول

✓ محاسبه نما در جدول توزیع فراوانی

$$Mo = a_i + \frac{d_1}{d_1 + d_2} \times c$$

d_1: اختلاف فراوانی گروه شامل نما با گروه ماقبل خود

d_2 : اختلاف فراوانی گروه شامل نما با گروه مابعد خود

✓ محاسبه میانه در جدول توزیع فراوانی

a_i : حد پایین دست اختلاف فراوانی گروه شامل نما با گروه ماقبل خود که میانه در آن قرار دارد.

c : فاصله دسته‌ها.

$$Md = a_i + \frac{\frac{N}{2} - F_{i-1}}{f_i} \times c$$

f_i : فراوانی دسته میانه.

F_{i-1} : فراوانی تجمعی دسته قبل از دسته میانه.

✓ محاسبه چارک ها در جدول توزیع فراوانی

$$Q_j = a_i + \frac{\frac{j \times N}{4} - F_{i-1}}{f_i} \times c$$

جدول توزیع فراوانی

جدول فراوانی داده تصویری زیر را تهیه نمایید.

105	221	183	186	121	181	180	143
97	154	153	174	120	168	167	141
245	228	174	199	181	158	176	110
163	131	154	115	160	208	158	133
207	180	190	193	194	133	156	123
134	178	76	167	184	135	229	146
218	157	101	171	165	172	158	169
199	151	142	163	145	171	148	158
150	175	149	87	160	237	150	135
196	201	200	176	150	170	118	149

جدول توزیع فراوانی

- فراوانی نسبی

$$f_i = \frac{n_i}{n}$$

- فراوانی تجمعی

$$F_i = F_{i-1} + f_i$$

$$F_0 = 0$$

فاصله دسته‌ها	فراوانی	فراوانی نسبی (%)	فراوانی تجمعی
70 to 90	2	2,50	2,50
90 to 110	3	3,75	6,25
110 to 130	6	7,50	13,75
130 to 150	14	17,50	31,25
150 to 170	22	27,50	58,75
170 to 190	17	21,25	80,00
190 to 210	10	12,50	92,50
210 to 230	4	5,00	97,50
230 to 250	2	2,50	100,00
Total	80	100	-

مثال نما یا مد

X	F
54 -57	1
57 -60	3
60 -63	6
63 -66	8
66 -69	2

← طبقه نما

$$MO = L + i\left(\frac{d_1}{d_1 + d_2}\right)$$

$$d_1 = 8 - 6 = 2$$

$$d_2 = 8 - 2 = 6$$

$$MO = 63 + 3\left(\frac{2}{2+6}\right) = 63/75$$

مثال میانه

x	f	CF
40-45	8	50
35 -40	7	42
30-35	5	35
25 -30	6	30
20-25	10	24
15 -20	8	14
10-15	6	6

$\Sigma f = 50$

طبقه میانه ←

$$m = L + \left(\frac{\frac{N}{2} - CF}{f}\right) \times i$$

$$m = 25 + \left(\frac{25-24}{6}\right)5 = 25/83$$

مثال

در توزیع فراوانی درس آمار یک کلاس، نمرات به شرح ذیل می‌باشد جدول فراوانی مربوط به توزیع را تهیه کنید؟

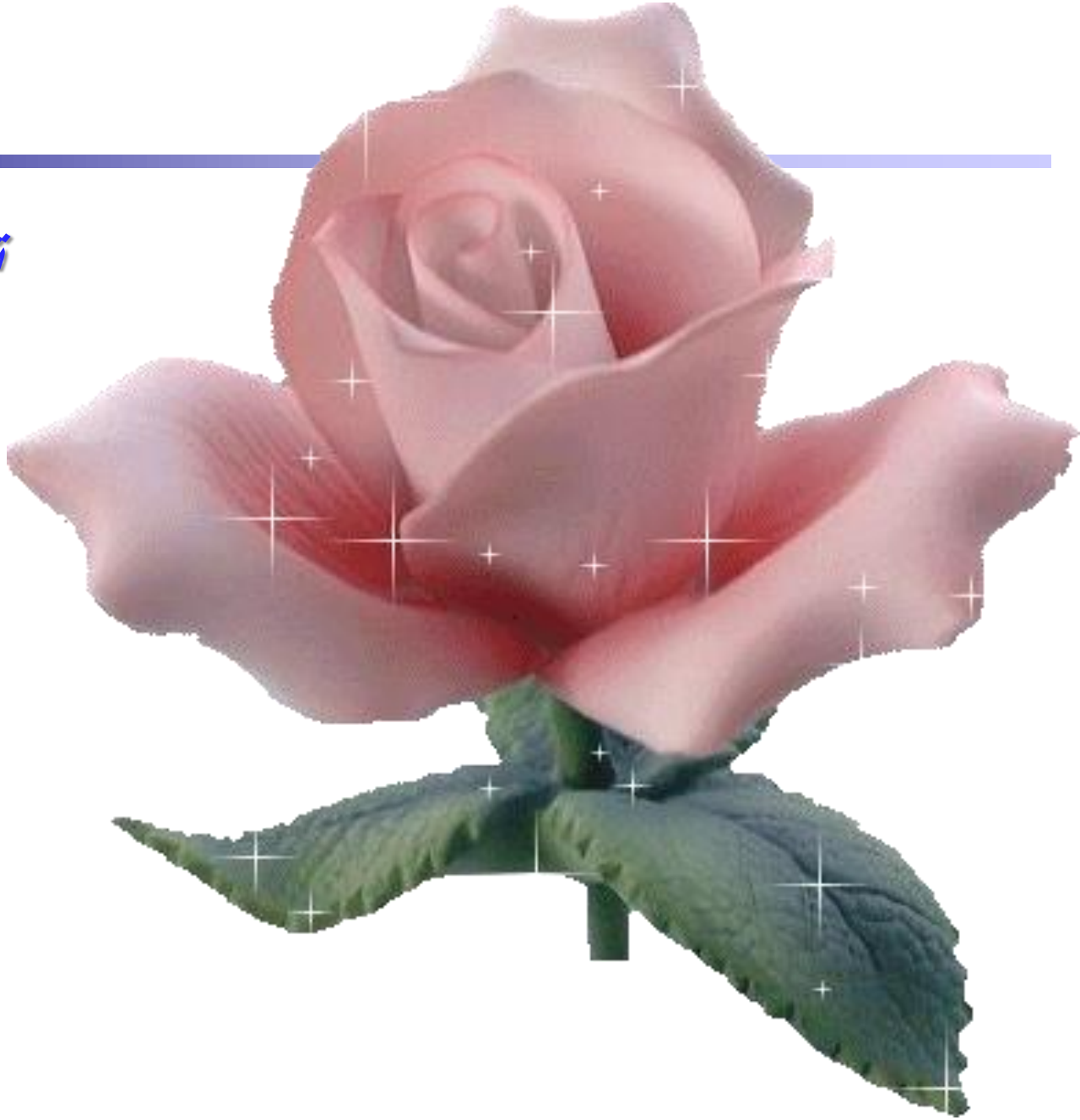
۱۰-۱۲-۱۱-۱۰-۱۱-۱۲-۱۰-۱۳-۱۵-۱۰

X	f
15	1
13	1
12	2
11	2
10	4

جواب:

با توجه به جدول فوق، عدد ۴ در ستون f بیانگر اینست که عدد ۱۰ چهار بار تکرار شده است.

نکته: اگر داده‌های ستون فراوانی (f) را با هم جمع کنیم تعداد کل داده ها بدست می آید .

یعنی $N = \sum f_i$ **در مثال فوق N = ۱۰**

تشکر از توجه شما