www.ainbpr.blogfa.com

آمار توصیفی

تعریف علم آمار: علم آمار به مجموعه روش‌های علمی اطلاق می‌شود که جهت جمع‌آوری اطلاعات اولیه، مرتب و خلاصه کردن، طبقه‌بندی و تجزیه و تحلیل اطلاعات اولیه استفاده می‌شود.

جامعه آماری: هر مجموعه از اشیاء که لااقل دارای یک صفت مشخص مشترک باشند جامعه آماری می‌نامند مانند جامعه افراد یک کلاس یا محصولات تولیدشده یک کارخانه

صفت مشخصه: صفتی است که بین تمام افراد جامعه مشترک می‌باشد مانند دانشجویان یک دانشگاه و.....

نمونه: به تعداد محدودی از اعضای جامعه آماری که بیان‌کننده تمام ویژگی‌های جامعه اصلی است نمونه می‌گویند که نمونه باید تصادفی باشد.

اگر صفتی در کلّ جامعه مورد اندازه‌گیری قرار گیرد به آن پارامتر و اگر در نمونه اندازه‌گیری شود به آن آماره می‌گویند.

معمولاً پارامترها در جامعه ثابت هستند ولی مجهول و از طریق نمونه‌گیری تصادفی و تجزیه و تحلیل اطلاعات نمونه و بدست آوردن مقدار آماره، مقدار پارامتر را تخمین می‌زنیم.

پارامترها و آماره‌های مهم

گروه	شاخص	نماد کلی	میانگین	واریانس	نسبت پیروزی
جامعه	پارامتر	θ	μ	6^2	P
نمونه	آماره	$\hat{\theta}$	$\bar{X}$	S^2	$\hat{P}$

آمار توصیفی: آن قسمت از آمار که درباره توصیف یک مجموعه از مشاهدات بحث می‌کند و هدف آن محاسبه پارامترهای جامعه است

آمار استنباطی: آن قسمت از آمار که با استفاده از تجزیه و تحلیل اطلاعات حاصل از نمونه به برآورد پارامترهای جامعه می‌پردازد.

آمار ناپارامتریک : در امار پارامتریک مشاهدات از توزیع نرمال برخوردارند ولی در آمار ناپارامتریک چنین نیست و بیشتر متغیرها با مقیاس کیفی سنجیده می‌شوند و آزاد از توزیع هستند.

صفات
- ثابت: بین تمام افراد یکسان است مانند دانشجو بودن
- متغیر: بین افراد جامعه از فردی به فرد دیگر تغییر می‌کند مانند سن، وزن و....

صفات متغیر
- کیفی: اسمی و ترتیبی
- کمی
  - گسسته ← قابل اندازه‌گیری است و بین مقادیر مختلف فاصله وجود دارد مانند تعداد افراد خانواده یا جمعیت شهرها
  - پیوسته ← مقادیر خودشان را از مجموعه اعداد حقیقی می‌گیرند مانند وزن، قد

www.ainbpr.blogfa.com

PDF created with pdfFactory Pro trial version www.pdffactory.com

صفات کیفی : صفاتی که واحد نداشته و قابل اندازه‌گیری نیستند ولی می‌توان آنها را طبقه‌بندی کرد مانند جنس، مرغوبیت کالا، گروه خونی و.....

صفات کمی : صفاتی که واحد داشته و قابل اندازه‌گیری و شمارش هستند مانند قد، وزن و.....

با توجه به نوع صفات کیفی و کمی مقیاس‌های متفاوتی برای اندازه‌گیری متغیرها وجود دارد مانند اسمی، رتبه‌ای، فاصله‌ای و نسبی

۱) **مقیاس اسمی:** هر کدام از حالات نسبت به دیگری مزیتی ندارند مانندگروه خونی و اهل کدام شهر بودن و....

۲) **مقیاس رتبه‌ای (ترتیبی):** بین حالات مختلف اولویت‌بندی وجود دارد و ترتیب در آنها اهمیت دارد مانند بررسی مرغوبیت چندکالا برحسب درجه ۱، ۲، ۳ یا قوی، متوسط، ضعیف یا دکترا، فوق لیسانس، لیسانس، کاردانی

۱) **مقیاس فاصله‌ای :** حالت ترتیبی رعایت شده و همچنین بین هر دو صفت فاصله مشخص وجود دارد و همچنین نقطه‌ای به عنوان صفر قراردادی در نظر گرفته می‌شود مانند درجه سانتیگراد

۲) **مقیاس نسبی :** مانند فاصله‌ایست ولی دارای صفر مطلق می‌باشد مانند گرم

انواع روش‌های نمونه‌گیری :

- **تصادفی ساده :** تمام افراد شانسی مساوی دارند و نمونه‌گیری یا به‌صورت قرعه‌کشی یا جدول اعداد تصادفی انجام می‌شود.
- **منظم:** از نقطه معینی شروع کرده و بقیه اعضاء نمونه را به ترتیب انتخاب می‌کنیم مانند شروع از یک شماره دانشجویی خاص
- **گروهی:** جامعه را به گروه‌های مختلفی تقسیم کرده سپس به صورت تصادفی ساده و منظم نمونه‌گیری می‌کنیم مانند بررسی عملکرد واحدهای مختلف یک سازمان . . . .
- **خوشه‌ای:** اگر جامعه بسیار گسترده باشد ابتدا گروه‌هایی از آن را انتخاب کرده و سپس در داخل آنها نمونه‌گیری می‌کنیم (هر دو حالت به‌صورت تصادفی یا منظم) مانند بررسی میزان تحصیل کارمندان یک کشور
- **مرحله‌ای:** حالت توسعه یافته نمونه‌گیری خوشه‌ای است. کشور ← چند شهر ← چند سازمان ← چند اداره ← چنداحد نمونه‌گیری می‌کنیم.(بررسی میزان تحصیل کارمندان)

حجم جامعه: تعداد افراد بررسی شده جامعه آماری که آن را با N نشان می‌دهیم.

نکاتی در مورد $\sum$: برای نشان دادن مجموع چند جمله از $\sum$ استفاده می‌کنیم:

$$x_1 + x_2 + \ldots + x_n = \sum_{i=1}^{n} x_i$$

$$\sum_{i=3}^{7} x_i = x_3 + x_4 + x_5 + x_6 + x_7$$

PDF created with pdfFactory Pro trial version www.pdffactory.com

www.ainbpr.blogfa.com

آمار

$$\sum_{i=1}^{5} 4 = 4+4+4+4+4 = 20$$

$$\sum (x_i + y_i) = \sum x_i + \sum y_i$$

$$\sum_{i=1}^{n} C x_i = C \sum_{i=1}^{n} x_i$$

نکته : اگر n داده $x_1 x_2 \dots x_n$ را در اختیار داشته باشیم به $\sum X_i$، مجموع داده‌ها و به $\sum X_i^2$ مجموع مجذور داده‌ها می‌گویند.

توزیع فراوانی و جدول فراوانی :

داده‌های کیفی یا کمی گسسته : اگر یک صفت، با حالت‌های $A_1, A_2, \dots A_k$ دارای مقادیر $f_1, f_2, \dots f_k$ باشد می‌توانیم جدول زیر را تشکیل دهیم

$$\sum_{i=1}^{k} f_i = N \quad \text{حجم جامعه}$$

صفت	f_i
A_1	f_1
A_2	f_2
⋮	⋮
A_k	f_k

کیفی مانند میزان تحصیلات 55 نفر از کارمندان یک شرکت .

صفت	فراوانی مطلق f_i
دیپلم	15
کاردانی	15
لیسانس	20
فوق‌لیسانس	5
	55

یا کمی گسسته مانند تعداد فرزندان 105 خانواده، به‌صورت 0 ، 1 ، 2 ، 3 ، 4 ، 5

صفت	f_i
0	5
1	10
2	15
3	20
4	25
5	40
	105

www.ainbpr.blogfa.com

PDF created with pdfFactory Pro trial version www.pdffactory.com

جدول فراوانی برای داده‌های کمی پیوسته :

اگر صفت مورد بررسی یک صفت پیوسته باشد مانند وزن افراد، طول، قد و... نمی‌توان برای هر نقطه خاص، فراوانی تعیین نمود. مثلاً بگوییم 5 نفر دارای وزن 70 کیلوگرم هستند چون وزن این افراد در فاصله 69.5 الی 70.5 می‌باشد در این صورت می‌توان داده‌ها را در طبقه‌بندی مشخصی در نظر بگیریم.

مراحل طبقه‌بندی :

۱- مرتب کردن داده‌ها به‌صورت صعودی

۲- تعیین دامنه تغییرات $R = X_{max} - X_{min} + 1$

۳- تعیین تعداد طبقات K: یا به‌صورت نظری با $K = 1 + 3.3 \log_{10} N$ یا $2^{k-1} \le N \le 2^k$

۴- تعیین فاصله طبقات L: $L = \frac{R}{k}$

جدول فراوانی می‌تواند شامل همه یا تعدادی از ستون‌های زیر باشد:

۱- فراوانی مطلق f_i : تعداد دفعات تکرار هر داده یا تعداد داده‌های مربوط به یک طبقه خاص $\sum_{i=1}^{k} f_i = n$

۲- فراوانی نسبی f_p $f_p = \frac{\text{فراوانی مطلق}}{\text{حجم کل نمونه}} = \frac{f_i}{n}$ $\sum_{i=1}^{k} f_p = 1$

۳- درصد فراوانی نسبی f'_p، $f'_p = f_p \times 100$ حاصل‌ضرب فراوانی نسبی در 100 $\sum_{i=1}^{k} f'_{p_i} = 100$

۴- فراوانی تجمعی F_i: فراوانی تجمعی هر طبقه برابر است با جمع فراوانی مطلق همان طبقه و طبقات ماقبل

$F_i = f_1 + f_2 + \ldots + f_i$

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

۵- فراوانی تجمعی نسبی F_p: مجموع فراوانی نسبی هر طبقه و طبقات ماقبل آن یا حاصل تقسیم فراوانی تجمعی یک طبقه بر تعداد کل داده‌ها

$F_{p_k} = f_{p_1} + f_{p_2} + \ldots + f_{p_k} = F_{p_{k-1}} + f_{p_k}$ یا $F_{p_i} = \frac{F_i}{n}$

۶- درصد فراوانی تجمعی نسبی F'_p: حاصل‌ضرب فراوانی تجمعی نسبی هر طبقه در 100 یا مجموع درصد فراوانی نسبی آن طبقه با طبقات ماقبل

۷- مرکز (نماینده) طبقه

$$\text{مرکز هر طبقه} = \frac{\text{حد پایین} + \text{حد بالا}}{2}$$

نکته : دقت نمایید که در حل بیشتر تست‌ها حدود طبقات می‌بایستی به صورت پیوسته باشد.

مثال) وزن 20 نفر از کارمندان یک کارخانه به‌صورت زیر اندازه‌گیری شده اگر تعداد طبقات 5 باشد جدول فراوانی را رسم نمایید

61-63-55-52-51-53-57-52-54-57-58-57-60-59-61-59-62-63-65-62

$R = X_{max} - X_{min} + 1 = 65 - 51 + 1 = 15$ طول طبقه $L = \frac{R}{K} = \frac{15}{5} = 3$

PDF created with pdfFactory Pro trial version www.pdffactory.com

www.ainbpr.blogfa.com

آمار

مرکز طبقه x_i	درصد فراوانی تجمعی نسبی	فراوانی تجمعی نسبی	فراوانی تجمعی F_i	درصد فراوانی نسبی	فراوانی نسبی	f_i	حدود طبقات	حدود واقعی
52	$\frac{4}{20}\times100$	$\frac{4}{20}$	4	$\frac{4}{20}\times100$	$\frac{4}{20}$	4	51-53	50.5 – 53.5
55	$\frac{6}{20}\times100$	$\frac{6}{20}$	6	$\frac{2}{20}\times100$	$\frac{2}{20}$	2	54-56	53.5 – 56.5
58	$\frac{10}{20}\times100$	$\frac{12}{20}$	12	$\frac{6}{20}\times100$	$\frac{6}{20}$	6	57-59	56.5 – 59.5
61	$\frac{17}{20}\times100$	$\frac{17}{20}$	17	$\frac{5}{20}\times100$	$\frac{5}{20}$	5	60-62	59.5 – 62.5
64	$\frac{20}{20}\times100=100$	$\frac{20}{20}=1$	$20=N$	$\frac{3}{20}\times100$	$\frac{3}{20}$	3	63-65	62.5 – 65.5
					1	$N=20$		

جهت مقایسه چند جامعه با یکدیگر از مشخص‌کننده‌های عددی استفاده می‌شود که عبارتند از :

۱- پارامترهای مرکزی

۲- پارامترهای پراکندگی

۳- پارامترهای انحراف از قرینگی (پارامترهای نسبی پراکندگی)

در حالت کلی دو نوع داده داریم

- داده‌های طبقه‌بندی نشده (مجزا ـ دسته بندی شده)
- داده‌های طبقه‌بندی شده

f_i ـ مقدار عددی داده x_i فراوانی مطلق

$x_i - f_i$ مرکز هر طبقه k تعداد طبقات فراوانی هر طبقه

* شاخص‌های مرکزی نشان می‌دهد که بیشتر افراد در یک جامعه آماری در کجا متمرکز شده‌اند ولی شاخص‌های پراکندگی نشان می‌دهد که افراد نسبت به یک مبدأ دلخواه تا چه اندازه پراکنده می‌باشند.

www.Ainbpr.blogfa.com

PDF created with pdfFactory Pro trial version www.pdffactory.com

www.ainbpr.blogfa.com

پارامترهای مرکزی:

۱- میانگین:

الف) میانگین حسابی

تعدادی از مشاهدات که بیشتر داده‌ها حول آن متمرکز شده‌اند میانگین حسابی نامیده می‌شود

$\mu = \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ در داده‌های مجزا

$\bar{X} = \frac{\sum_{i=1}^{k} f_i x_i}{n}$ در داده‌های طبقه‌بندی شده ، x_i : مرکز دسته

مثال : میانگین داده‌های 3,4,5,8,10 برابر است با $\bar{X} = \frac{\sum X_i}{5} = \frac{30}{5} = 6$

مثال :

طبقات	x_i	f_i
3-7	5	3
8-12	10	2
13-17	15	1
18-22	20	5

$\bar{X} = \frac{1}{11}(15+20+15+100) \cong 13.5$

x_i	f_i
2	5
3	10
4	2
5	3

$\bar{X} = \frac{\sum f_i x_i}{\sum f_i} = \frac{10+30+8+15}{20} = \frac{63}{20}$

نکته : به‌طور پیش‌فرض منظور از میانگین، میانگین حسابی می‌باشد چون میانگین هندسی و میانگین هارمونیک در شرایط خاصی استفاده می‌شوند.

مثال: اگر میانگین $x_1 x_2 \ldots x_{10}$ برابر 5 باشد و عدد 10 به آنها اضافه شود میانگین 10 و $x_1, x_2, \ldots, x_{10}$ چیست؟

$$5 = \frac{\sum_{i=1}^{10} x_i}{10} \rightarrow \sum_{i=1}^{10} x_i = 50 \Rightarrow \frac{\sum_{i=1}^{10} X_i + 10}{11} = \frac{50+10}{11} = \frac{60}{11}$$

مثال : میانگین داده‌های زیر را بدست آورید:

12345 , 12346 , 12348 , 12349

$X_i = 5, 6, 8, 9$ مقدار ثابت $= 12340$, $E(X_i) = \mu_{x_i} = \frac{5+6+8+9}{4} = 7$

$E(X_i + 12340) = E(X_i) + 12340 = 7 + 12340 = 12347$

ب) میانگین هندسی: میانگین هندسی معمولاً برای بدست آوردن مقدار متوسط نسبت‌ها، درصدها، شاخص‌ها و نرخ رشدها مورد استفاده قرار می‌گیرد و در جایی مورد استفاده خواهد بود که تمام داده‌ها مثبت باشند.

PDF created with pdfFactory Pro trial version www.pdffactory.com

www.ainbpr.blogfa.com

$\bar{X}_G = \sqrt[n]{x_1 \times x_2 \times \dots x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i}$ داده‌های مجزا

$\bar{X}_G = \sqrt[n]{x_1^{f_1} \times x_2^{f_2} \times \dots x_k^{f_k}} = \sqrt[n]{\prod_{i=1}^{k} x_i^{f_i}}$ طبقه‌بندی شده

مثال: میزان تولید کارخانه‌ای در 4 سال متوالی 6,4,2 و 27 برابر شده است (نسبت به سال قبل) مطلوب است محاسبه میزان افزایش متوسط تولید ؟

$\bar{X}_G = \sqrt[4]{2 \times 4 \times 6 \times 27} = \sqrt[4]{8 \times 6 \times 3^3} = \sqrt[4]{2^3 \times 3^3 \times 6} = \sqrt[4]{6^3 \times 6} = \sqrt[4]{6^4} = 6$

مثال: تعداد کارکنان کارخانه‌ای طی 4 سال متوالی 180,190 , 160 , 150 بوده است. متوسط رشد سالانه کارکنان چنددرصد است؟

$\bar{X}_G = \sqrt[3]{\frac{160}{150} \times \frac{180}{160} \times \frac{190}{180}} = \sqrt[3]{\frac{190}{150}} = 1.082$ $\quad 1.082 - 1.000 = 0.082 \times 100 = 8.2\%$

فرمول کلی : $\sqrt[\text{۱ ـ سال}]{\frac{\text{مقدار سال آخر}}{\text{مقدار سال اول}}}$

ج) میانگین هارمونیک ـ همساز ـ توافقی: اگر تمام داده‌ها مخالف صفر باشند و کار ثابتی با سرعت‌های متفاوت انجام شود و محاسبه متوسط سرعت کار مورد نظر باشند از میانگین همساز استفاده می‌کنیم.

$\bar{X}_H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} \rightarrow \frac{1}{H} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{X_i}$ داده‌های مجزا

- یعنی عکس میانگین همساز عبارت است از میانگین حسابی عکس داده‌ها

$\bar{X}_H = \frac{n}{\sum_{i-1}^{k} \frac{f_i}{x_i}}$ داده‌های طبقه‌بندی شده

از میانگین همساز معمولاً در مواردی استفاده می‌شود که می‌خواهیم میانگین نسبت‌هایی را محاسبه کنیم که صورت و مخرج آنها دارای مقیاس یکسان نیستند ولی صورت نسبت‌ها مساوی می‌باشند. اگر مخرج نسبت‌ها مساوی باشند از میانگین حسابی استفاده می‌کنیم) (معمولاً برای محاسبه حد متوسط سرعت‌ها استفاده می‌شود)

مثال: اتومبیلی فاصله بین دو شهر A, B را که شصت کیلومتر است با سرعت متوسط 60 $\frac{km}{h}$ رفته و همین مسیر را با سرعت متوسط 30 $\frac{km}{h}$ برگشته است. سرعت متوسط رفت و برگشت چقدر است؟ $\quad n=2 \quad V_1 = \frac{x_1}{t} = \frac{60}{t}, V_2 = \frac{x_2}{t} = \frac{60}{t}$

$V_1 = 60 \quad V_2 = 30$

$H = \frac{2}{\frac{1}{30} + \frac{1}{60}} = 40$

www.ainbpr.blogfa.com

PDF created with pdfFactory Pro trial version www.pdffactory.com

www.ainbpr.blogfa.com

مثال : 2 اتومبیل مسافتی را با سرعت 100 رفته هنگام برگشت اولی با سرعت 80 بر می‌گردد و دومی نصف مسیر را با سرعت 60 و مابقی را با سرعت 40 برمی‌گردد. متوسط سرعت کدام است؟

$$\bar{X}_H = \frac{2+1+\frac{1}{2}+\frac{1}{2}}{\frac{2}{100}+\frac{1}{80}+\frac{\frac{1}{2}}{60}+\frac{\frac{1}{2}}{40}}$$

مثال: اتومبیلی یک مسیر 200 کیلومتری را با سرعت 80 کیلومتر بر ساعت رفته هنگام برگشت 40 کیلومتر را با سرعت 100 کیلومتر بر ساعت و مابقی را با سرعت 120 کیلومتر بر ساعت برمی‌گردد سرعت متوسط چقدر است؟

$$X_H = \frac{1+\frac{1}{5}+\frac{4}{5}}{\frac{1}{80}+\frac{\frac{1}{5}}{100}+\frac{\frac{4}{5}}{120}}$$

مثال) اتومبیلی مسیری را با سرعت 100 کیلومتر رفته، $\frac{1}{3}$ مسیر را با سرعت km80 و باقیمانده را با سرعت km120 برگشته است. سرعت متوسط اتومبیل چقدر است؟

$$\bar{X}_H = \frac{\sum w_i}{\sum \frac{w_i}{x_i}} = \frac{1+\frac{1}{3}+\frac{2}{3}}{\frac{1}{100}+\frac{\frac{1}{3}}{80}+\frac{\frac{2}{3}}{120}} = 101.4$$

- اگر داده‌ها دارای وزن باشند داریم:

$$\bar{X}_w = \frac{\sum_{i=1}^{n} X_i w_i}{\sum_{i=1}^{n} w_i}$$

د) میانگین وزنی: برای محاسبه معدل دانشجویان استفاده می‌شود.

ه) میانگین درجه دوم:

$Q = \sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2}$ داده‌های مجزا

$Q = \sqrt{\frac{1}{n}\sum_{i=1}^{k} f_i x_i^2}$ داده‌های طبقه‌بندی شده

- نکته مهم $\bar{X}_H \le \bar{X}_G \le \bar{X} \le Q$

خواص مهم میانگین:

۱ــ مجموع انحرافات داده‌ها از میانگین صفر است

$$\sum (X_i - \bar{X}) = 0 \quad \text{یا} \quad \sum (X_i - \mu) = 0$$

www.ainbpr.blogfa.com
PDF created with pdfFactory Pro trial version www.pdffactory.com

موسسه آموزش عالی آزاد پارسه
www.ainbpr.blogfa.com

آمار

$$\sum_{i=1}^{k} f_i(x_i-\bar{x})=0$$

- $\sum (x_i-a)^2$ به ازای $a=\mu$ کمترین مقدار خود را دارد.

- اگر به تمام داده‌ها مقدار ثابت a را اضافه کنیم یا از تمام داده‌ها مقدار ثابت a را کم کنیم. میانگین به اندازه a اضافه یا کم می‌شود.

$$y_i=x_i\pm a \Rightarrow \bar{y}=\bar{x}\pm a$$

- اگر تمام داده‌ها را در عدد ثابت b ضرب کنیم میانگین نیز در b ضرب می‌شود.

$$y_i=b\,x_i \Rightarrow \bar{y}=b\bar{x}$$

در حالت کلی $y_i=bx_i+a \Rightarrow \bar{y}=b\bar{x}+a$

- میانگین تنها پارامتری است که اگر به جای کلیه داده‌ها قرار گیرد مجموع آنها تغییری نمی‌کند.

1,2,3 $\bar{X}=2$ ، 2,2,2

مجموع 6 مجموع=6

مثال: در شرکتی میانگین حقوق افراد 120 هزار تومان است مطلوب است میانگین حقوق افراد در هر یک از حالات زیر :

الف) به حقوق هر شخص 10000 تومان اضافه و کم شود.

ب) حقوق هر شخص را نصف کنیم.

ج) از حقوق هر شخص 20% حقوقش کم شود.

د) به حقوق هر شخص 20% میانگین اضافه شود.

$$E(X+10)=E(X)+10=130000$$

$$E(X-10)=E(X)-10=110000$$

$$E\left(\frac{X}{2}\right)=\frac{1}{2}E(X)=60$$

$$E(X-0.2X)=E(0.8X)=0.8E(X)=0.8\times 120000=96000$$

$$E(X+0.2(\mu))=E(X+0.2(120000))=E(X+24000)=E(X)+24000=144000$$

۲- میانه: (M_d): عددی است که 50% داده‌ها قبل از آن و 50% داده‌ها بعد از آن قرار می‌گیرند.

اگر n (تعداد داده‌ها) فردباشد $\text{میانه}=x_{\frac{n+1}{2}}$ ابتدا به صورت صعودی مرتب می‌کنیم ← داده‌های مجزا

اگر n (تعداد داده‌ها)زوج باشد $\text{میانه}=\frac{x_{\left(\frac{n}{2}\right)}+x_{\left(\frac{n}{2}+1\right)}}{2}$ میانگین دو عدد وسطی

مثال:

1 2 3 4 5 $M_d=3$

1 2 3 4 5 6 $M_d=\frac{3+4}{2}=3.5$

www.ainbpr.blogfa.com
PDF created with pdfFactory Pro trial version www.pdffactory.com

www.ainbpr.blogfa.com

داده‌های مجزا ولی دسته‌بندی شده

x_z = میانه اگر مقدار فوق عدد صحیحی بود مانند z ← محل میانه $\frac{N}{2}+\frac{1}{2}$

اگر مقدار فوق اعشاری بود داریم جزء اعشاری :d جزء صحیح: z

میانه $= x_z + d(x_{z+1} - x_z)$

مثال

X_i	0	1	2	3
f_i	10	20	30	40
F_i	10	30	60	

محل میانه $\frac{N}{2}+\frac{1}{2} = 50.5$

مقدار میانه $= 2 + 0.5(2-2) = 2$

مثال:

X_i	0	1	2	3
f_i	10	40	20	30
F_i	10	50		

محل میانه $= \frac{N}{2}+\frac{1}{2} = 50.5$

مقدار میانه $= 1 + 0.5(2-1) = 1.5$

داده‌های طبقه‌بندی شده:

ابتدا دسته شامل میانه را تعیین می‌کنیم. (اولین دسته‌ای که فراوانی تجمعی آن بزرگتر یا مساوی نصف تعداد داده‌ها است) و سپس داریم:

$$M_d = L + \frac{\frac{n}{2} - F_{M-1}}{f_m} \times w$$

L: کران پایین دسته میانه

F_{M-1}: فراوانی تجمعی طبقه ماقبل میانه f_m: فراوانی مطلق طبقه میانه w: طول هر طبقه

طبقات	f_i	F_i
10-20	10	10
20-30	20	30
30-40	30	60
40-50	40	100

$$M_d = 30 + \frac{50-30}{30}(10) = 36.66$$

خواص میانه:

۱ - میانه از اعداد بسیار بزرگ و کوچک متأثر نمی‌شود.

۲ - مجموع قدرمطلق انحرافات تک تک داده‌ها از میانه همیشه کمترین مقدار است یا به عبارتی $\sum |X_i - a|$ به ازای $a = M_d$ کمترین مقدار خود را خواهد داشت.

www.ainbpr.blogfa.com
PDF created with pdfFactory Pro trial version www.pdffactory.com

۳- مد یا نما M_o

متغیری که دارای بیشترین فراوانی است. مثلاً وقتی که می‌گوییم مد نمرات یک کلاس آمار 17 است یعنی بیشترین تعداد دانشجویان 17 گرفته‌اند.

داده‌های مجزا: داده‌ای که فراوانی مطلق آن از سایر داده‌ها بیشتر است.

مثال: 1,1,3,2,2 ($M_o=1,2$) 1,1,2,2 مد نداریم 2 3 3 4 4 4 5 ($m_o=4$)

داده‌های طبقه‌بندی نشده (دسته‌بندی شده)

X_i	0	1	2	3
f_i	40	20	50	12

$\rightarrow M_o=2$

داده‌های طبقه‌بندی شده:

۱) ابتدا طبقه شامل مد را پیدا می‌کنیم (طبقه‌ای که بیشترین فراوانی مطلق را دارد) و سپس داریم،

$$M_o=L+\frac{d_1}{d_1+d_2}w$$

L : کران پایین دسته مد دار

W : طول هر طبقه

d_1 : فراوانی مطلق طبقه ماقبل مد – فراوانی مطلق طبقه مددار

d_2 : فراوانی مطلق طبقه مابعد مد – فراوانی مطلق طبقه مددار

طبقات	f_i
80 -100	20
100 -120	35
120 -140	40
140 -160	5

$d_1=40-35=5$

$d_2=40-5=35$

$M_o=120+\frac{5}{5+35}20=122.5$

۴- چندک‌ها:

چندک‌ها مقادیری از مشاهدات هستند که دامنه تغییرات را به فواصلی تقسیم می‌کنند که فراوانی‌ها در هر یک از این فواصل، درصد معینی از فراوانی کل را تشکیل می‌دهد. چندک p ام یعنی p ٪داده‌ها در سمت چپ آن (کوچک‌تر از آن) و (1−p)% داده‌ها بزرگتر از آن هستند مثلاً چارک اول یعنی مقداری که 25٪ داده‌ها کوچک‌تر از آن و 75٪ داده‌ها بزرگتر از آن هستند.

مثلاً وقتی می‌گوییم دهک دوم حقوق کارمندان یک شرکت 60 هزار تومان است یعنی حقوق 20% از کارمندان کمتر یا مساوی 60 هزار تومان بوده و حقوق 80% کارمندان بیشتر از 60 هزار تومان است.

PDF created with pdfFactory Pro trial version www.pdffactory.com

محاسبه چندک در داده‌های مجزا: ابتدا داده‌ها را به صورت صعودی مرتب می‌کنیم و مقدار $np+\frac{1}{2}$ را محاسبه کرده و اگر عدد صحیحی بود $X_{np+\frac{1}{2}}$ محل چندک را نشان می‌دهد. در صورتی‌که $\left(np+\frac{1}{2}\right)$ مقدار صحیحی نباشد جزء صحیح آن را z و جزء اعشاری را با d نمایش می‌دهیم و خواهیم داشت.

$$Q_p = x_z + d\left(x_{z+1} - x_z\right) = (1-d)x_z + d\,x_{z+1}$$

چارک $p=\frac{1}{4}\ \frac{2}{4}\ \frac{3}{4}$ دهک $p=0.1\ 0.2\ \ldots 0.9$ صدک $p=0.01, 0.02, \ldots, 0.99$

مثال: در داده‌های روبه رو دهک چهارم چقدر است؟

$-1\ \ 4\ \ 0\ \ 2\ \ 6\ \ 9 \rightarrow -1\ \ 0\ \ 2\ \ 4\ \ 6\ \ 9$

$n=6 \quad p=0.4$

$$np+\frac{1}{2} = 6(0.4)+0.5 = 2.9$$

$$z=2,\ d=0.9 \qquad Q_{0.4} = x_z + d\left(x_{z+1}-x_z\right) = x_2 + 0.9\left(x_3 - x_2\right) = 0 + 0.9(2-0) = 1.8$$

نکته : ممکن است بعضی از چندک‌ها روی هم منطبق شوند مثلاً دهک دوم مساوی صدک 20 ام می‌باشد یا چارک سوم مساوی صدک 75 ام می‌باشد تنها نقطه‌ای که چارک بر روی دهک منطبق می‌شود نقطه میانه است

نکته : چارک دوم = میانه = دهک پنجم = صدک پنجاه

محاسبه چندک در داده‌های طبقه‌بندی شده :

ابتدا دسته شامل چندک را مشخص می‌کنیم.(اولین طبقه‌ای که فراوانی تجمعی آن بیشتر یا مساوی np باشد). سپس داریم:

$$\text{مقدار چندک} = L + \frac{np - F_{i-1}}{f_i} \times W$$

F_{i-1}: فراوانی تجمعی طبقه ماقبل چندک

f_i: فراوانی مطلق طبقه چندک

W: طول طبقه

L : کران پایین طبقه چندک

	f_i	F_i
40 – 50	5	5
50 – 60	18	23
60 – 70	7	30
		n=30

$30\times 0.2 = 6$

مثال: مطلوب است دهک دوم

$$Q_{0.2} = 50 + \frac{6-5}{18}(10) = 50.55$$

نکته : اگر طبقات گسسته بود حتماً باید حدود طبقات را تبدیل به پیوسته کنیم

نکته : هرگاه دو چندک متوالی با هم برابر باشند هر چندکی که بین آنها باشد با آن دو چندک برابر است زیرا تمام داده‌های ما بین آن دو چندک با یک دیگر برابر و مساوی مقدار آن دو چندک هستند $Q_1 = 10 \ldots Q_3 = 10 \Rightarrow Q_2 = 10$

www.ainbpr.blogfa.com
PDF created with pdfFactory Pro trial version www.pdffactory.com

پارامترهای پراکندگی

میزان پراکندگی داده‌ها نسبت به یکدیگر یا نسبت به میانگین را مورد بررسی قرار می‌دهند.

الف) دامنه تغییرات : بی‌ثبات‌ترین پارامتر پراکندگی است چون تابع دو عدد است. $R = X_{max} - X_{min}$

$1, 4, 7, 1000 \rightarrow R = 1000 - 1 = 999$

$1, 944, 997, 1000 \rightarrow R = 1000 - 1 = 999$

ب) دامنه میان چارکی $IQR = Q_3 - Q_1$ ، IQR یا 25% بالای میانه تا 25% پایین میانه

انحراف چارکی $SIQR = \frac{IQR}{2}$

در توزیع‌های نامتقارن از میانه به عنوان شاخص مرکزی و از انحراف چارکی به عنوان شاخص پراکندگی استفاده می‌شود.

نکته : در توزیع‌های نامتقارن یا توزیع‌هایی که حدود طبقات اول و آخر نامشخص است بهترین پارامتر مرکزی میانه و بهترین پارامتر پراکندگی انحراف چارکی است.

ج) انحراف متوسط از میانگین:

$$A.D = \frac{1}{N}\sum_{i=1}^{n} |X_i - \mu| \Rightarrow$$

د) واریانس و انحراف معیار: میانگین مربع انحراف داده‌ها از میانگین،واریانس نامیده می‌شود. جذر واریانس برابر انحراف معیار است .

بهترین شاخص برای نشان دادن انحرافات بزرگ می‌باشد یعنی بیش از سایر پارامترها تحت تاثیر انحرافات بزرگ قرار دارد.

- واریانس اعداد برابر، صفر خواهد شد

- واریانس اعداد 1800 ، 1800 ، ...و 1800 چند است؟ صفر

واریانس جامعه $\sigma_x^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{x})^2}{n} = \frac{\sum X_i^2}{n} - \bar{X}^2$

$$\sigma_x^2 = \frac{\sum f_i (x_i - \bar{x})^2}{n} = \frac{\sum f_i x_i^2}{n} - \left(\frac{\sum f_i x_i}{n}\right)^2$$

نکته : در صورتی که در صورت مسأله کلمه نمونه ذکر شود واریانس به طریق دیگری محاسبه می‌شود.

واریانس نمونه $S_x^2 = \frac{\sum (X_i - \bar{x})^2}{n-1} = \frac{\sum X_i^2 - n\bar{X}^2}{n-1}$

$\sigma_{x \pm a} = \sigma_x \qquad \sigma_{bx} = b\sigma_x$

- اگر به تک تک داده‌ها مقدار ثابت a را اضافه یا کم کنیم واریانس داده‌های جدید با واریانس داده‌های قبلی برابر است

PDF created with pdfFactory Pro trial version www.pdffactory.com

www.ainbpr.blogfa.com

- اگر تک‌تک داده‌ها را در عدد ثابت a ضرب کنیم واریانس داده‌های جدید a^2 برابر واریانس داده‌های قدیم خواهد بود

نکته: برای داده‌های N.....3 2 1 داریم: امید ریاضی $\frac{N(N+1)}{2N}=\frac{N+1}{2}$ واریانس $\frac{N^2-1}{12}$

میانگین چند جامعه $\mu=\frac{N_1\mu_1+N_2\mu_2+...N_k\mu_k}{N_1+N_2+...+N_k}$

ه) ضریب تغییرات CV

$CV=\frac{\sigma}{\bar{X}}$ یا CV درصد $\frac{\sigma}{\bar{X}}\times 100$

در مقایسه دو جامعه با میانگین و واحد اندازه‌گیری یکسان، هر جامعه‌ای که دارای واریانس یا انحراف‌معیار کمتری داشته باشد بهتر است.

دقت بیشتر و خطای کمتر و تمرکز بیشتر و پراکندگی کمتر ← انحراف معیار (واریانس کمتر) $\sigma_1^2<\sigma_2^2 \rightarrow \sigma_1<\sigma_2$

برای مقایسه چند جامعه با یکدیگر در شرایط زیر از CV استفاده می‌کنیم :

- دو یا چند جامعه دارای واحدهای اندازه‌گیری مختلفی باشند
- صفات یکسان بوده ولی بزرگی مشاهدات به طور قابل ملاحظه‌ای تفاوت دارد.
- چند جامعه با میانگین‌های متفاوت

مثال : اگر داشته باشیم $\sum_{i=1}^{3} X_i=3$ ، $\sum_{i=1}^{3} X_i^2=6$ ، $CV=?$

$$\sigma^2=\frac{\sum X_i^2}{n}-\bar{X}^2=\frac{6}{3}-(1)^2=1\rightarrow\sigma=1 \quad \bar{X}=1\Rightarrow CV=\frac{\sigma}{\bar{X}}=\frac{1}{1}=1$$

مثال: میانگین و انحراف معیار حقوق در یک سازمان به ترتیب 50 هزار تومان و 20 هزار تومان است. اگر حقوق‌ها در این سازمان 25% افزایش یابند ضریب تغییرات حقوق چه خواهد شد؟

$$E(X+0.25x)=E(1.25x)=1.25E(x)$$
$$\sigma(x+0.25x)=\sigma(1.25x)=1.25\sigma_x$$

$$CV=\frac{\sigma}{\bar{x}}=\frac{1.25\sigma_x}{1.25E(x)}=\frac{\sigma}{\bar{X}}$$

بنابراین ضریب تغییرات تغییری نخواهد کرد

چولگی :

چولگی به معنی عدم تقارن توزیع می‌باشد. در صورتی که بین نمودارهای مختلف به دنبال میزان عدم متقارن یا عدم قرینگی باشیم می‌توانیم از معیاری به نام چولگی استفاده کنیم

توزیع‌ها را می‌توان به سه دسته تقسیم نمود.

www.ainbpr.blogfa.com

PDF created with pdfFactory Pro trial version www.pdffactory.com

۱) متقارن : در توزیع متقارن داده‌های میانی بیشترین فراوانی را دارند و در دو نیمه پایین و بالای توزیع، فراوانی‌ها به یک نسبت پراکنده‌اند $sk=0$ در توزیع متقارن میانه و مد و میانگین روی هم و در یک نقطه قرار می‌گیرند. $M_o = M_d = \bar{X}$

۲- چوله به راست: داده‌های دارای بزرگترین فراوانی‌ها در بین داده‌های پایین قرار دارند و داریم $M_o < M_d < \bar{X}$

$sk>0$ (توزیع متمایل به چپ) (چولگی مثبت)

۳- چوله به چپ: داده‌های دارای بزرگ‌ترین فراوانی در بین داده‌های بالا قرار دارند. (چولگی منفی) $sk<0$ (توزیع متمایل به راست)

$\bar{X} < md < M_o$

* دقت نمایید که همیشه میانه بین میانگین و مد قرار دارد.

ضریب چولگی $sk = \dfrac{\mu_3}{\sigma^3} = \dfrac{\text{گشتاور مرکزی مرتبه سوم حول میانگین}}{(\text{انحراف معیار})^3} = \dfrac{\frac{\sum(X_i-\mu)^2}{N}}{\sigma^3}$

$sk_1 = \dfrac{\mu - M_o}{\sigma}$

$sk_2 = \dfrac{3(\mu - M_d)}{\sigma}$

نکته : در توزیع‌هایی که چولگی زیاد نباشد رابطه تقریبی زیر به نام رابطه تقریبی کارل پیرسن بین مقادیر میانگین، میانه و نما وجود دارد:

(میانه ـ میانگین) $3 \cong$ مد ـ میانگین

$\bar{X} - M_o \cong 3(\bar{X} - md)$

www.ainbpr.blogfa.com

PDF created with pdfFactory Pro trial version www.pdffactory.com

مثال : در یک توزیع با چولگی خفیف میانگین 20 و مد 8 است. مقدار میانه چقدر است؟

$$(20-8)=3(20-M_d) \Rightarrow M_d=16$$

نکته : تفسیر قدرمطلق ضریب چولگی :

هر چه قدرمطلق ضریب چولگی بزرگتر باشد تفاوت جامعه از نظر قرینگی با توزیع نرمال بیشتر خواهد بود:

تقریباً چولگی وجود ندارد ـ جامعه تقریباً نرمال $|sk|\leq 0.1$

چولگی خیلی کم ـ جامعه تفاوت کمی با نرمال دارد $0.1<|sk|\leq 0.5$

چولگی زیاد ـ جامعه تفاوت زیادی با نرمال دارد $|sk|>0.5$

مثال : اگر $N=10$ و $\Sigma(X_i-\mu_x)^2=40$ ، $\Sigma(X_i-\mu_x)^3=80$ باشد مقدار ضریب چولگی چقدر است؟

$$sk=\frac{\mu_3}{\sigma^3}=\frac{\sum\frac{(X_i-\mu)^3}{N}}{\sigma^3}=\frac{\frac{80}{10}}{2^3}=\frac{8}{8}=1$$

$$\sigma^2=\frac{\sum(X_i-\mu_x)^2}{x}=\frac{40}{10}=4 \rightarrow \sigma=2$$

کشیدگی و ضریب کشیدگی:

کشیدگی معیاری است بدون واحد که ارتفاع منحنی را مورد بحث قرار می‌دهد

$$\text{میزان کشیدگی هر توزیع دلخواه}=\frac{\mu_4}{\sigma^4}=\frac{\frac{\sum(x_i-\mu)^4}{N}}{(\text{واریانس})^2}$$

کشیدگی توزیع نرمال همیشه ۳ است

معیار ضریب کشیدگی اختلاف کشیدگی توزیع‌ها را نسبت به کشیدگی توزیع نرمال بدست می‌آورد.

$$\text{ضریب کشیدگی } E=\frac{\mu_4}{\sigma^4}-3$$

تحلیل ضریب کشیدگی :

$E=0$ کشیدگی توزیع هم ارتفاع نرمال است

$E>0$ کشیدگی توزیع از نرمال بلندتر و پراکندگی آن از نرمال کمتر است.

$E<0$ کشیدگی توزیع از نرمال کوتاه‌تر و پراکندگی آن از نرمال بیشتر است.

تفسیر قدرمطلق ضریب کشیدگی

$|E|\leq 0.1$: توزیع جامعه از جهت پراکندگی و کشیدگی تقریباً نرمال است (تفاوت قابل اغماض)

$0.1<|E|\leq 0.5$: توزیع جامعه از جهت پراکندگی و کشیدگی تفاوت اندکی با توزیع نرمال دارد (تفاوت غیرقابل اغماض)

PDF created with pdfFactory Pro trial version www.pdffactory.com

www.ainbpr.blogfa.com

$|E| > 0.5$: توزیع جامعه از جهت پراکندگی و کشیدگی اختلاف فاحشی با توزیع نرمال دارد (تفاوت غیرقابل اغماض)

گشتارو حول مبدا

$$m_k = \frac{\sum_{i=1}^{N} X_i^k}{N} = \frac{\sum f_i x_i^k}{N} \text{ یا } m_k = E(X^k)$$

نکته : گشتاور مرتبه اول حول مبدأ برابر میانگین است.

$$m_1 = \frac{\sum x_i}{N} = \bar{X} \text{ یا } m_2 = \frac{\sum X_i^2}{N} \text{ یا } m_3 = \frac{\sum X_i^3}{N}, \ldots$$

نکته : گشتاور مرتبه دوم حول مبدأ برابر $\frac{\sum X_i^2}{n}$ می‌باشد.

گشتاور حول میانگین یا گشتاور مرکزی :

$$\mu_k = \frac{\sum_{i=1}^{N} (X_i - \bar{X})^k}{N} = \frac{\sum f_i (x_i - \bar{x})^k}{N}$$

نکته : گشتاور مرتبه اول حول میانگین برابر صفر است

$$\mu_1 = \frac{\sum (X_i - \bar{X})}{N} = 0 \qquad \mu_2 = \frac{\sum (X_i - \bar{X})^2}{N} = \text{var}(X) \qquad \mu_3 = \frac{\sum (X_i - \bar{X})^3}{N} \qquad \mu_4 = \frac{\sum (X_i - \bar{X})^4}{N}$$

واریانس = گشتاور مرکزی مرتبه دوم حول میانگین

مثال اگر $\sum f_i (x_i - \mu)^4 = 5000, N = 1000$ و انحراف معیار جامعه 2 باشد مقدار ضریب کشیدگی کدام است؟

$$E = \frac{\mu_4}{\sigma^4} - 3 = \frac{\frac{5000}{1000}}{(2^2)^2} - 3 = \frac{5}{16} - 3 = -2.69$$

www.ainbpr.blogfa.com

PDF created with pdfFactory Pro trial version www.pdffactory.com

www.ainbpr.blogfa.com

نمودارهای فراوانی:

۱- **نمودارهای وصفی** : این دسته از نمودارها برای نمایش هندسی داده‌های کیفی بکار می‌روند مانند مشاهداتی از نوع خوب، بد، متوسط یا گروه A,B,C و....

الف) نمودار ستونی: این نمودار را در یک دستگاه مختصات که محور افقی آن نشان دهنده تعداد حالات مربوط به مشاهدات کیفی و محور عمودی آن نشان دهنده فراوانی مطلق یا نسبی هر گروه است رسم می‌کنند.

ب) نمودار دایره‌ای (مهم):

یکی از ابزارهای مناسب جهت نمایش داده‌های کیفی و تجسم مشاهدات است و متناظر با هر یک از حالات، فراوانی نسبی مربوطه را در 360 ضرب کرده و زاویه مورد نظر مربوط به آن حالت خاص را محاسبه می‌کنیم. سپس تمام زاویای بدست آمده مربوط به کل حالات را در دایره‌ای نشان می‌دهیم.

$$\text{زاویه مورد نظر } \alpha = \frac{f_i}{n} \times 360^\circ$$

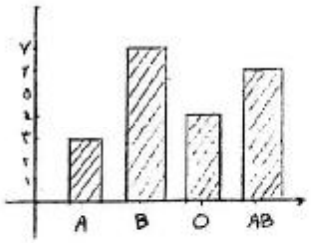
ج) نمودار پاره‌تو

نوعی نمودار ستونی برای داده‌های کیفی است. در این نمودار فراوانی هر موضوع روی محور عمودی و نوع آن روی محور افقی آورده می‌شود در نمودار پاره‌تو حالتی که دارای بیشترین فراوانی است در سمت چپ قرار می‌گیرد و به صورت نزولی بقیه حالات را رسم می‌کنیم این نمودار 3 محور دارد که محور سوم آن نشان دهنده فراوانی‌های نسبی تجمعی است و به صورت یک محور عمودی دیگر و در انتهای محور افقی آورده می‌شود. همچنین منحنی روی نمودار پاره‌تودرصدهای تجمعی k امین طبقه را به هم وصل می‌کند.

نمودار پاره تو در تحلیل موجودی انبار کالاها، نواقص سیستم‌ها، توزیع درآمد و توزیع کارمندان سازمان‌ها و کنترل کیفیت کاربرد فراوانی دارد.

مثال) فرض کنید در بررسی گروه خونی 20 کارمند جدول زیر بدست آمده است. نمودار ستونی و دایره‌ای آن را رسم کنید.

گروه خونی	f_i
A	3
B	7
O	4
AB	6
	20

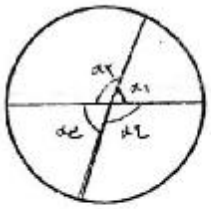


$$\alpha_i = \frac{f_i}{n} \times 360 \Rightarrow \alpha_1 = \frac{3}{20} \times 360 = 54^\circ$$

$$\alpha_2 = \frac{7}{20} \times 360 = 126^\circ$$

$$\alpha_3 = \frac{4}{20} \times 360 = 72^\circ$$

$$\alpha_4 = \frac{6}{20} \times 360 = 108^\circ$$



www.ainbpr.blogfa.com

PDF created with pdfFactory Pro trial version www.pdffactory.com

www.ainbpr.blogfa.com

۲- نمودارهای کمی:

الف) نمودار بافت‌نگار(هیستوگرام): یکی از نمودارهای مورد استفاده جهت نمایش داده‌های کمی پیوسته می‌باشد. در این نمودار محور افقی مربوط به حدود واقعی طبقات و محور قائم مربوط به فراوانی مطلق یا فراوانی نسبی می‌باشد. با توجه به حدود طبقات می‌توان حدود واقعی طبقات را بدست آورد.

ب) نمودار چندضلعی :

نمودار چندضلعی نموداری است که محور افقی مربوط به مرکز دسته‌ها (طبقه‌ها) و محور عمودی مربوط به فراوانی نسبی یا مطلق هر یک از طبقات مربوطه می‌باشد. پس از بدست آوردن نقاط مربوط به طبقات، به نقاط مزبور دو نقطه فرضی دیگر اضافه می‌کنیم، اولی مرکز طبقه ماقبل اولین طبقه و دیگری نمایندهٔ طبقه مابعد آخرین طبقه می‌باشد. در این حالت ابتدا و انتهای نمودار روی محور افقی قرار می‌گیرد از اتصال متوالی این نقاط به یکدیگر نموداری حاصل می‌شود که آن را نمودار چندضلعی می‌نامند.

در تحقیقاتی که شامل دو یا چند جامعه آماری هستند و هدف مقایسه آنهاست. ترجیح داده می‌شود که توزیع‌های فراوانی جوامع روی یک نمودار رسم گردد. در چنین حالتی برای سهولت در خواندن و درک اطلاعات از نمودارهای چندضلعی فراوانی استفاده می‌شود.

ج) نمودار فراوانی تجمعی :

توزیع فراوانی تجمعی توزیعی است که تعداد مشاهدات قبل از یک نقطه معین را در مقیاس مشاهدات نشان می‌دهد و جهت محاسبه چندک ها و مقایسه هندسی دو یا چند توزیع فراوانی به‌طور همزمان بکار می‌رود.

دو روش رسم برای رسم نمودار فراوانی تجمعی وجود دارد:

۱- محور افقی بر اساس متوسط طبقات مدرج می‌شود و محور عمودی براساس فراوانی‌های تجمعی طبقات به منحنی بدست آمده در این روش پلی‌گن فراوانی تجمعی می‌گویند.

۲- محور افقی عبارت از کرانه‌های بالای هر طبقه می‌باشد که به آن منحنی فراوانی تجمعی می‌گویند محور عمودی نیز بر اساس فراوانی‌های تجمعی طبقات می‌باشد.

از نمودار تجمعی برای پاسخ به سوالاتی همچون چه تعداد از بطری‌ها حاوی مایعی کمتر از A میلی‌لیتر هستند؟ نیز استفاده می‌شود برای یافتن پاسخ کافی است که خطی عمودی از A به نمودار وصل گردد. سپس با خواندن فراوانی تجمعی آن جواب مشخص خواهد شد.

این نمودار جهت مقایسه توزیع‌های فراوانی تجمعی دو یا چند جامعه که از نظر تعداد با هم مساوی هستند مفید خواهد بود مثلا برای مقایسه میزان رشد تورم در دو یا چند کشور می‌توان از این نمودار استفاده کرد. هر چه شباهت نمودارها بیشتر باشد شباهت رشد تورم در آن کشورها بیشتر است.

www.ainbpr.blogfa.com

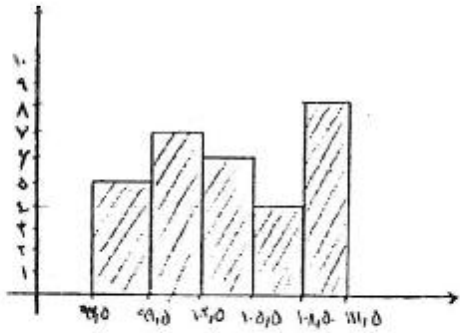
PDF created with pdfFactory Pro trial version www.pdffactory.com

www.ainbpr.blogfa.com

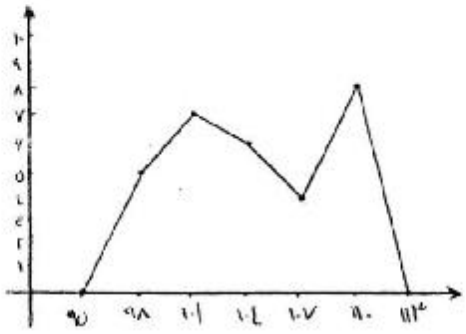
مثال) نمودارهای کمی مربوط به جدول فراوانی زیر رسم کنید.

حدود واقعی	مرکز دسته	فراوانی مطلق f_i	فراوانی تجمعی F_i
96.5 -99.5	98	5	5
99.5 -102.5	101	7	12
102.5 -105.5	104	6	18
105.5 – 108.5	107	4	22
108.5 - 111.5	110	8	30

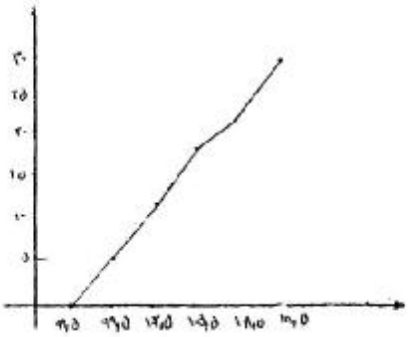
الف) نمودار بافت نگار



ب) نمودار چندضلعی فراوانی



ج) نمودار فراوانی تجمعی



www.ainbpr.blogfa.com

PDF created with pdfFactory Pro trial version www.pdffactory.com

www.ainbpr.blogfa.com

نمودارهای تحلیل اکتشافی داده‌ها:

یکی دیگر از نمودارهای کمی می‌باشند که اغلب در مراحل اولیه تحلیل داده‌ها مفید هستند. نمودار ریشه و برگ و نمودار جعبه‌ای از انواع تحلیل اکتشافی داده‌ها هستند.

برای تهیه نمودار شاخه و برگ ارقام مربوط به مشاهدات را به دو بخش تقسیم می‌کنند. شاخه شامل یک یا چند رقم اولیه و برگ که شامل ارقام باقی‌مانده است. مثلاً مقداری مثل 25 را به شاخه 2 و برگ 5 می‌توان تقسیم کرد. نمودار شاخه و برگ به ما این امکان را می‌دهد که بعضی از خصوصیات مشاهدات را که به شکل اولیه قابل شناسایی نیستند به سرعت تعیین کنیم مانند محاسبه مد، چندک‌ها و... . برخی مواقع ممکن است برگ یا شاخه بیشتری مورد نیاز باشد، تأمین این نیاز تعدیل شاخه‌های اصلی است.

نمودار جعبه‌ای نیز یکی از مفیدترین نمودارهای اکتشافی برای مقایسهٔ دو یا چند جامعه آماری است که نشان دهنده چارک‌ها و حداقل و حداکثر مشاهدات است. خطی که جعبه را به دو نیم تقسیم می‌کند میانه مشاهدات است و ابتدای جعبه، چارک اول و انتهای آن، چارک سوم است. از هر طرف جعبه نیز به اندازهٔ نقاط مرزی مشاهدات (حداقل و حداکثر) خطی ادامه می‌یابد که گاهی ریشه نامیده می‌شود.

www.ainbpr.blogfa.com

PDF created with pdfFactory Pro trial version www.pdffactory.com