- تعریف آمار:

در اصطلاح عامیانه آمار به معنای ثبت و نمایش اطلاعات عددی در باب موضوع اگر، مثلاً ثبت و نمایش میزان تخم‌مرغ تولیدی کشور، ترکیب تریلی، اما از نظر علمی، علم آمار عبارتی از یکپارچه‌ای است برای جمع‌آوری، تنظیم، خلاصه، تجزیه و تفسیر مشاهدات (داده‌ها)، تعیین اعتبار و تعمیم نتایج آنها کمک می‌کند.

۱- جمع‌آوری مشاهدات (Collection of data):

جمع‌آوری مشاهدات به دو روش نمونه‌گیری (Sampling) و سرشماری (census) صورت می‌پذیرد. به این اعمال در اصطلاح آماری تهیه می‌شود.

۲- تنظیم مشاهدات (organization of data):

شامل معرفی اندازه‌ها یا شمارش‌های جمع‌آوری شده در قالب مناسب به منظور استنتاج نتایج منطقی آن است. روش‌های تنظیم و معرفی مشاهدات توسط جداول و نمودارها صورت می‌پذیرد.

۳- تجزیه مشاهدات (analysis of data):

مرحله استخراج اطلاعات قابل اطمینان به صورت ارقام خلاصه شده و قابل درک از داده‌های جمع‌آوری شده است.

مهمترین شاخص‌هایی که برای این منظور به کار برده می‌شوند عبارتند از:

- شاخص‌های تمایل به مرکز

- شاخص‌های پراکندگی

- شاخص‌های شکل توزیع

۴- تفسیر مشاهدات (interpretation data):

مرحله نتیجه گیری از مشاهدات تجزیه شده آن. این نتیجه گیری به کل جامعه یا موارد مشابه تعمیم داده می شود.

تقسیم بندی آمار:

علم آمار را به دو روش آمار پارامتری و آمار ناپارامتری تقسیم می کنند. آمار پارامتری به آمار توصیفی و آمار استنباطی تقسیم می شود.

آمار توصیفی (descriptive statistic):

شامل آن بخش از آمار پارامتری است که به جمع آوری، خلاصه کردن، نمایش و پردازش داده های آماری می پردازد تا پارامترهای آماری و ویژگیهای طی داده های مزبور را به سادگی و روشنی بنمایاند. این نوع آمار صرفاً به توصیف جامعه می پردازد و هدف آن تا کجا پارامترهای جامعه است.

آمار استنباطی (inferential statistics):

در این بخش از آمار، اطلاعات بدست آمده از آمار توصیفی مورد تجزیه و تحلیل قرار می گیرد، براساس تحلیل های انجام شده نتیجه گیری و استنباط تحویل می آید و این نتیجه گیری به کل جامعه یا موارد مشابه تعمیم داده می شود. در این نوع آمار محقق با استفاده از مقادیر نمونه، آماره ها را می سنجد. سپس با کمک تخمین و آزمون فرض های آماری، آماره ها را به پارامترهای جامعه تعمیم می دهد.

مثال: فرض کنید آمار تخته خرده چوب (تخته خرده چوب) مربوط به چند کارخانه تولید تخته خرده چوب کشور را در اختیار باشد، اگر قرار باشد در مورد متوسط تولید این کارخانه ها گزارشی تهیه شود، روشهای آمار توصیفی مورد استفاده قرار می گیرند، حال اگر این اطلاعات را برای برآورد تولید کل تخته خرده چوب در کشور یا میزان تولید سال آینده مورد استفاده قرار می گیرند وارد بحث استنباطی شده ایم

رابطه بین آمار توصیفی و آمار استنباطی:

تئوری احتمالات به ما توانایی آن را می دهد تا از آمار توصیفی برآمار استنباطی استدلال می سازد، زیرا این استدلالها درباره نمونه را به طور قاطع یک بار و صد اطمینان به شخص می توان به جامعه تعمیم داد. احتمال، ابزاری برای سنجش و اندازه گیری موارد اطمینان و عدم اطمینان است.

آمار ناپارامتریک:

این نوع آمار در مقابل آمار پارامتریک بیان می شود. فرض اساسی در آمار پارامتریک برخوردار بودن مشاهدات از توزیع نرمال است در حالیکه در آمار ناپارامتریک این فرض ضروری ندارد. آزمونهایی که در آمار ناپارامتریک بکار می روند را آزمونهای بدون توزیع نرمال گویند.

اصولا بخش عظیمی از پژوهشهای علمی با تغیرهای کیفی سنجیده می شوند که فاقد توزیع آماری هستند و آنها را از آزمایشهای توزیع نرمال گویند، در اینگونه موارد بهتر است از روشهای ناپارامتریک استفاده شود.

جامعه آماری:

جامعه آماری مجموعه ای از افراد یا عناصر است که دست کم دارای یک خصوصیت (صفت) مشترک باشند. در این تعریف توجه داشته

ذیل مهم است:

- منظور از فرد یا عضو واحدی است که اندازه گیری می شود.
- تعداد عناصر هر جامعه آماری را حجم یا اندازه آن جامعه گویند و با حرف N نشان می دهند.
- هر صفت یا خصوصیت مشترک که بین همه عناصر (افراد) جامعه آماری وجود دارد، جامعه را طوری که آن صفت مشخص می کند

جامعه‌ی متناهی :

جامعه‌ای آماری که از نظر تعداد افراد (عناصر) محدود باشد مثل : ارتفاع درختان کاج چند ساله‌ی شهرستان ذ ، اعضای جامعه

جنگلهای ایران ، ... .

جامعه‌ی نامتناهی :

جامعه‌ای که تعداد افراد آن نامحدود یا آنقدر زیاد و غیرقابل شمارش باشد، هرچه از افراد این جامعه برداریم، تمام نمی‌شود مثل:

تعداد نقاط واقع بر یک صفحه ، تعداد اعداد حقیقی بین صفر و یک و ... .

نمونه :

به هر بخشی از جامعه ، معرف آن باشد نمونه گفته می‌شود . منظور از معرف بودن آن یک ، همه صفات جامعه باید در نمونه وجود داشته باشد.

صفت مهم از نظر تحقیق باید تناسب در نمونه و جامعه داشته باشد.

نمونه‌ها را به دو دسته اریب (biased) و تصادفی (random) تقسیم می‌کنند. نمونه‌ای را که آمارگر در انتخاب آن تمایل و سلیقه‌ی آماری دخالت دارد، نمونه‌ی اریب گویند. در نمونه‌های تصادفی همه عناصر جامعه شانس یکسانی جهت انتخاب شدن به عنوان عضوی از نمونه دارند. بر این اساس در علم آمار نمونه‌گیری اریب ارزش چندانی ندارند و منظور از نمونه ، همان نمونه تصادفی است.

تعداد عناصر نمونه را اندازه یا حجم نمونه گویند و با n نشان می‌دهند. اگر $n \geq 30$ باشد نمونه را بزرگ و اگر $n < 30$ باشد نمونه کوچک است.

شاخص‌های آماری:

عبارتست از مقادیر (کمیت‌های) معیارهایی، برای نشان دادن نتیجه یک بررسی آماری به کار برده می‌شوند. به شاخص‌های آماری در جامعه، پارامتر (Parameter) و در نمونه، آماره (Statistic) گفته می‌شود.

پارامتر:

پارامتر به صفت یا خصوصیتی از کل جامعه گفته می‌شود قابل اندازه‌گیری باشد. به عنوان مثال گلو متوسط درآمد کارکنان دولت یا استخوان از اندازه‌گیری درآمد تمامی کارکنان دولت حاصل می‌شود به آن پارامتر گویند.

آماره:

می‌دانیم، هدف‌های تحقیق، تعیین پارامترهای جامعه است. اندازه‌گیری صفت یا خصوصیت تک تک عناصر جامعه

بعلت نیاز به هزینه و وقت زیاد، فرضی بودن برخی از جوامع و عدم دسترسی به کلیه عناصر جامعه مشکل و یا غیرممکن است

بنابراین می‌توان با گرفتن نمونه‌گیری برای مقدار پارامتر برآوردی ارائه داد (از طریق اندازه‌گیری صفت یا خصوصیت در نمونه انتخابی) حاصل بدست آمده از این نمونه‌گیری را به برآوردی از پارامتر آن آماره گویند. البته به دلیل خطای نمونه‌گیری ممکن است آماره محاسبه شده، برآوردی دقیق از پارامتر نباشد.

معمولاً پارامترها را با حروف یونانی و آماره‌ها را با حروف اول انگلیسی نشان می‌دهند

شاخص	پارامتر	آماره
میانگین	μ	$\bar{X}$
واریانس	σ^2	S^2
انحراف از معیار	σ	S
نسبت	π	P
همبستگی	ρ	r
تعداد مشاهدات	N	n

داده‌های آماری :

در یک بررسی آماری، مجموعه اطلاعاتی که در مورد موضوع مورد مطالعه را داده‌های آماری می‌گویند. به دو دسته تقسیم می‌شوند:

۱- داده‌های کمی (عددی) : قابل اندازه‌گیری بوده و به صورت عدد نوشته می‌شوند، مانند قد، وزن، سن، مقاومت و ...

۲- داده‌های کیفی (توصیفی) : قابل اندازه‌گیری کمی نبوده و به صورت غیرعددی نشان داده می‌شوند مانند گروه خون، شغل، رنگ چشم و ...

صفات و متغیرها :

تغیر کمیت یا کیفیتی می‌باشد که مقدار آن از یک عضو جامعه آماری به عضو دیگر جامعه آماری متفاوت است.

واژه‌های پژوهشگران آمار، صفت (character) نامیده می‌شود. در واقع

اندازه‌گیری از کوچکترین واحد نمونه‌گیری را داده گویند و اندازه هر یک را داده یا متغیر می‌نامند، زیرا مقدار آنها در افراد جامعه

آماری متفاوت است. بررسی آماری همواره پیرامون صفت متغیر و یا X نشان داده می‌شود صورت می‌پذیرد.

متغیرها را می‌توان به کمی و کیفی تقسیم نمود (مثلاً نوعی داده مثل رنگ) متغیرهای کیفی ترتیبی اسمی

متغیر یک آمار یا پیوسته است (Discrete) و یا پیوسته (Continuous) باشد.

- متغیر ناپیوسته: متغیری است که فقط مجموعه‌ای از ارزشهای معین را آن اختصاص داده می‌شود، در اینجا متغیرها، ارزشهای

مجاور بین دو مقدار یا دو ارزش دارای معنی و مفهوم نیست. اینگونه متغیرها هم به صورت کیفی هستند (مقاطع تحصیلی و جنس) و هم به صورت کمی

(مانند تعداد دانشجویان درس آمار و یا تعداد تولید کارخانه صنعت سلیمان) هستند.

- متغیر پیوسته: متغیری است که هر ارزش یا مقدار (اعشاری و کسری) را می‌توان به آن اختصاص داد و هر عددی که بین دو

واحد آن انتخاب می‌شود دارای معنی و مفهوم است مثل وزن محصول تکه، ارتفاع درخت، ... اینگونه متغیر قاعدتاً کمی است.

Subject:

date:

آمار توصیفی

جدول توزیع فراوانی ۱

برای استخراج اطلاعات نخفته در داده‌های تحقیق (داده‌های خام) که به صورت سرشماری (مشاهده) و یا اندازه‌گیری به ما می‌آیند، لازم است در ابتدا داده‌ها را خلاصه، طبقه‌بندی، مرتب و تنظیم نمود تا قابل تجزیه و تحلیل و نتیجه‌گیری گردند. این کار به وسیله جدول توزیع فراوانی صورت می‌پذیرد.

یک جدول توزیع فراوانی شامل ستون‌های ذیل است:

۱- فراوانی (f_i): تعداد داده‌ها در هر طبقه از جدول توزیع فراوانی را فراوانی آن طبقه می‌گوییم و با f_i (فراوانی طبقه i ام جدول) نشان می‌دهیم

$f_1 + f_2 + \cdots + f_k = n$ ← تعداد کل داده‌ها

f_1: فراوانی طبقه ۱

f_k: فراوانی طبقه k ام

$n = \sum_{i=1}^{k} f_i \Rightarrow 1 \leq f_i \leq k$

۲- فراوانی نسبی (f_{ri}): اگر f_i را بر تعداد کل داده‌ها تقسیم کنیم، فراوانی نسبی طبقه i ام یا f_{ri} به دست می‌آید

$f_{ri} = f_i / n$

$f_{r_1} + f_{r_2} + \cdots + f_{r_k} = 1$, $\sum_{i=1}^{k} f_{ri} = 1 \Rightarrow 0 \leq f_{ri} \leq 1$

۳- فراوانی تجمعی (f_{ci}): مجموع فراوانی‌های طبقه ۱ تا طبقه i ام را با f_{ci} نشان داده و فراوانی تجمعی طبقه i ام گوییم.

$f_{c_1} = f_1$, $f_{c_2} = f_1 + f_2$, $f_{ci} = f_1 + f_2 + \cdots + f_i$

$f_{ck} = n$

۴- فراوانی تجمعی نسبی (f_{cri}): مجموع فراوانی‌های نسبی طبقه ۱ تا طبقه i ام را با f_{cri} نشان داده و فراوانی تجمعی نسبی طبقه i ام گوییم.

$f_{cr_1} = f_{r_1}$, $f_{cr_2} = f_{r_1} + f_{r_2}$

$f_{cr_i} = f_{r_1} + f_{r_2} + \cdots + f_{ri}$, $f_{cr_k} = 1$

۵- فراوانی درصد تراکمی نسبی (f_{cp_i}): حاصلضرب فراوانی تراکمی نسبی در عدد ۱۰۰ ← $f_{cp_i} = f_{cri} \times 100$

$\sum_{i=1}^{k} f_{cp_i} = 100$

phoenix
Scanned with CamScanner

Subject:

date:

برحسب نوع داده‌ها (گسسته یا پیوسته) نحوه تشکیل جدول توزیع فراوانی متفاوت است که در حوصله این درس خارج می‌باشد

فقط جهت آشنایی جدول توزیع فراوانی داده‌های پیوسته ارایه می‌شود:

مثال: وزن‌های 40 طالب‌آموز که به نزدیک‌ترین عدد صحیح گرد شده‌اند به شکل زیر است، یک جدول توزیع فراوانی برای این داده‌ها تشکیل دهید.

52	35	24	47	36	51	34	38	46	33
47	36	38	50	47	34	41	40	42	40
26	29	30	32	30	35	37	37	41	21
31	30	26	35	45	23	43	31	34	43

1) $K = 1 + 3.322 \log n = ?$ (تعداد طبقات) $\quad K = 1 + 3.322 \log 40 = 6.321 \simeq 7$

2) S (میزان تغییرپذیری در داده‌ها) $= \frac{\text{واحد گروه‌بندی داده‌ها}}{2} = 1/2 = 0.5$

3)
$$\begin{cases} X_{max} = \text{بزرگترین داده} + S = 52 + 0.5 = 52.5 \\ X_{min} = \text{کوچکترین داده} - S = 21 - 0.5 = 20.5 \end{cases}$$

$$R = X_{max} - X_{min} = 52.5 - 20.5 = 32$$

(طول طبقات) C یا $W = R/K = 32/7 = 4.5714 \simeq 5$

حدود طبقات	حدودشان (مرز طبقه نام)	x_{ci} (مرکز طبقه)	f_i	f_{ri}	f_{ci}	f_{cri}
20.5 - 25.5		23	3	0.075	3	0.075
25.5 - 30.5		28	6	0.15	9	0.225
30.5 - 35.5		33	10	0.25	19	0.475
35.5 - 40.5		38	8	0.20	27	0.675
40.5 - 45.5		43	6	0.15	33	0.825
45.5 - 50.5		48	5	0.125	38	0.95
50.5 - 55.5		53	2	0.05	40	1
			40	1		

phoenix
Scanned with CamScanner

Subject:

date:

بعد از تنظیم و مرتب کردن داده ها در آمار توصیفی در مرحله بعد لازم است که برای به دست آوردن نتایج مقدماتی درباره جمعیت مورد مطالعه . در یک نمودار ، داده ها را در یک یا چند عدد خلاصه کنیم . به چنین اعدادی شاخص یا معیار گوئیم و به ۳ دسته تقسیم می شوند : ۱- شاخص های تمایل به مرکز ۲- شاخص های پراکندگی ۳- شاخص های شکل توزیع

در این بحث فقط در زمینه ۱ شاخص اول صحبت می شود .

۱- شاخص های تمایل به مرکز :

شاخص های تمایل به مرکز، مقادیری هستند که عموما میزان تجمع داده ها حول یک مرکز داده ها را در یک مختص توزیع داده ها را نشان می دهند

به ۵ دسته تقسیم می شوند :

۱- میانگین حسابی

۲- میانه

۳- نما

۴- میانگین هندسی

۵- میانگین هارمونیک

۱- میانگین حسابی : بهترین شاخص تمایل به مرکزی باشد :

در جمعیت داده ها : μ = میانگین جامعه $= \frac{\sum x_i}{N}$

در نمونه : $\bar{X}$ = میانگین نمونه $= \frac{\sum x_i}{n}$

در جدول توزیع فراوانی :

$\bar{x} = 1/n \sum_{i=1}^{k} f_i x_i$ → جدول توزیع فراوانی داده های گسسته

$\bar{x} = 1/n \sum_{i=1}^{k} f_i x c_i$ → جدول توزیع فراوانی داده های پیوسته

phoenix
Scanned with CamScanner

Subject:

date:

مثال :

حدود طبقات	f_i	x_{ci}	$f_i x_{ci}$
1.45-1.95	4	1.7	6.8
1.95-2.45	6	2.2	13.2
2.45-2.95	20	2.7	54
2.95-3.45	7	3.2	22.4
3.45-3.95	10	3.7	37
3.95-4.45	3	4.2	12.6
	50		146

$\Rightarrow \bar{x} = \frac{1}{n} \sum_{i=1}^{6} f_i x_{ci} = \frac{146}{50} = 2.92$

۲- میانه : در یک مجموعه داده‌های مرتب شده آماری، میانه عددی است که 50٪ داده‌ها از آن کوچکتر و 50٪ دیگر داده‌ها از آن بزرگتر می‌باشد.

در مجموعه داده‌های مرتب شده آماری :

اگر n عدد فرد باشد $\Rightarrow$

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

اگر n عدد زوج باشد $\Rightarrow$

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

مثال : مطلوبست محاسبه میانه :

14 ، 11 ، 3 ، 7 ، 9 ، 11 ، 7 ، 5 ، 4

$\hookrightarrow$ 3 ، 4 ، 5 ، 7 ، 7 ، 9 ، 11 ، 11 ، 14 $\Rightarrow Md = X_5 = 7$

12 ، 7 ، 3 ، 4 ، 3 ، 5

3 ، 3 ، 4 ، 5 ، 7 ، 12 $\Rightarrow Md = \frac{X_3 + X_4}{2} = 4.5$

در جدول توزیع فراوانی :

در ابتدا لازم است تا طبقه میانه را پیدا کنیم :

طبقه میانه : اولین طبقه‌ای است در جدول توزیع فراوانی که فراوانی تراکمی آن مساوی و یا بزرگتر از نصف داده‌ها باشد

$$M_d = L_{Md} + \left(\frac{\frac{n}{2} - f_{c_{Md-1}}}{f_{Md}} \right) \times C$$

L_{Md} : حد پایین طبقه میانه

$f_{c_{Md-1}}$: فراوانی تراکمی طبقه قبل از طبقه میانه

f_{Md} : فراوانی طبقه میانه

C : طول طبقات

Scanned with CamScanner

Subject:

date:

مثال :

حدود طبقات	f_i	f_{ci}
17.25 - 20.35	2	2
20.35 - 23.45	7	9
23.45 - 26.55	10	19
26.55 - 29.65	17	36 → طبقه میانه
29.65 - 32.75	3	39
32.75 - 35.85	6	45
35.85 - 38.95	5	50
	$n=50$	

$$Md = 26.55 + \left(\frac{25-19}{17}\right) \times 3.1$$

$$Md = 27.644$$

۳- نهایاب : کمیتی که در معیاری تمایل به مرکز است و عبارت است از داده‌ای که فراوانی آن از بقیه داده‌ها بیشتر باشد. در مجموعه داده‌های مربوط به آماری، می‌تواند بیش از یک نهایاب وجود داشته باشد و یا در صورتی که داده‌ها دارای فراوانی یکسانی باشند اصلاً نهایاب وجود نداشته باشد

مثال : در مجموعه داده‌های ذیل نهایاب را بدست آورید.

الف) 3, 5, 7, 4, 7, 6 $\Rightarrow Mo = 7$

ب) 3, 5, 5, 4, 7, 7, 6, 6 $Mo_1 = 5 \quad Mo_2 = 6 \quad Mo_3 = 7$

ج) 3, 5, 4, 7, 6 → نهایاب اصلاً وجود ندارد.

در جدول توزیع فراوانی :

در ابتدا لازم است تا طبقه نهایی تعریف شود. طبقه نهایی طبقه‌ای است در جدول که بیشترین فراوانی را دارا می‌باشد.

حدود طبقات	f_i
99.5 - 109.5	3
109.5 - 119.5	5
119.5 - 129.5	4
129.5 - 139.5	7 → طبقه نهایی
139.5 - 149.5	6

L_{Mo} = حد پایین طبقه نهایی

d_1 = تفاضل فراوانی طبقه نهایی از طبقه ماقبل

d_2 = تفاضل فراوانی طبقه نهایی از طبقه بعد

C = طول طبقات

$$Mo = L_{Mo} + \left(\frac{d_1}{d_1 + d_2}\right) \times C$$

$$Mo = 129.5 + \left(\frac{3}{3+1}\right) \times 10$$

$$Mo = 137$$

phoenix
Scanned with CamScanner

Subject:

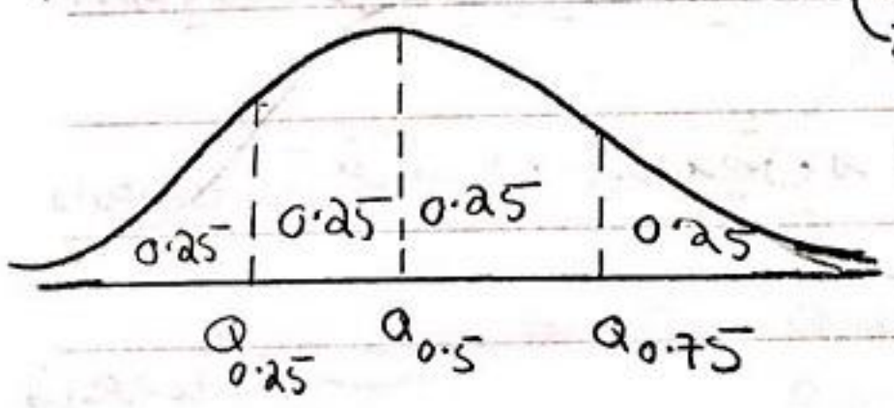
date:

چندک ها (صدک ها) :

چندک ها مقادیری هستند که مجموعه داده های مرتب شده آماری را به فاصله ای خیلی نزدیک تقسیم می کنند. چنین تقسیم بندی وضعیت نسبی هر یک از مشاهدات را در کل توزیع به ما نشان می دهد. چندک ها را با نماد Q_p نشان می دهند. اندیس p نشان دهنده نوع چندک است. p هم عددی است که از صفر تا یک p مشاهدات از آن کمتر می باشد. p همین صفر تا یک.

انواع چندک ها :

۱- چارک ها: مشاهدات را به 4 قسمت مساوی تقسیم می کند $p=0.25, 0.5, 0.75$. چارک اول = $Q_{0.25}$: مقداری است که 25/ داده ها از آن کوچکتر و 75/ از آن بزرگتر می باشند. چارک دوم = $Q_{0.5}$ ← میانه ، چارک سوم = $Q_{0.75}$



۲- دهک ها: مجموعه داده های مرتب شده را به ۱۰ قسمت مساوی تقسیم می کند $p=0.1, 0.2, \dots, 0.9$

$Q_{0.1}$ ← دهک دوم ، $Q_{0.2}$ ... $Q_{0.9}$ ← دهک نهم

$Q_{0.5}$ ← دهک پنجم = میانه

۳- صدک ها: مجموعه داده های مرتب شده آماری را به ۱۰۰ قسمت مساوی تقسیم می کند. $p=0.01, 0.02, \dots, 0.99$

$Q_{0.01}$ ← صدک دوم ، $Q_{0.02}$ ... $Q_{0.99}$ ← صدک نود و نهم

$Q_{0.50}$ ← صدک پنجاهم ← میانه

محاسبه چندک ها در جدول توزیع فراوانی :

تعداد داده ها

طبقه چندک : اولین طبقه ای است در جدول که فراوانی تجمعی آن مساوی یا بزرگتر از $n \cdot p$ باشد $f_{c_i} \geq n \cdot p$ یا $n \cdot p$ ← نوع چندک

فراوانی تجمعی طبقه قبل از طبقه چندک

طول طبقات ← C

$$Q_p = L_p + \left(\frac{nP - f_{c_{p-1}}}{f_p} \right) \times C$$

حد پایین طبقه چندک

فراوانی طبقه چندک

Scanned with CamScanner

Subject:

date:

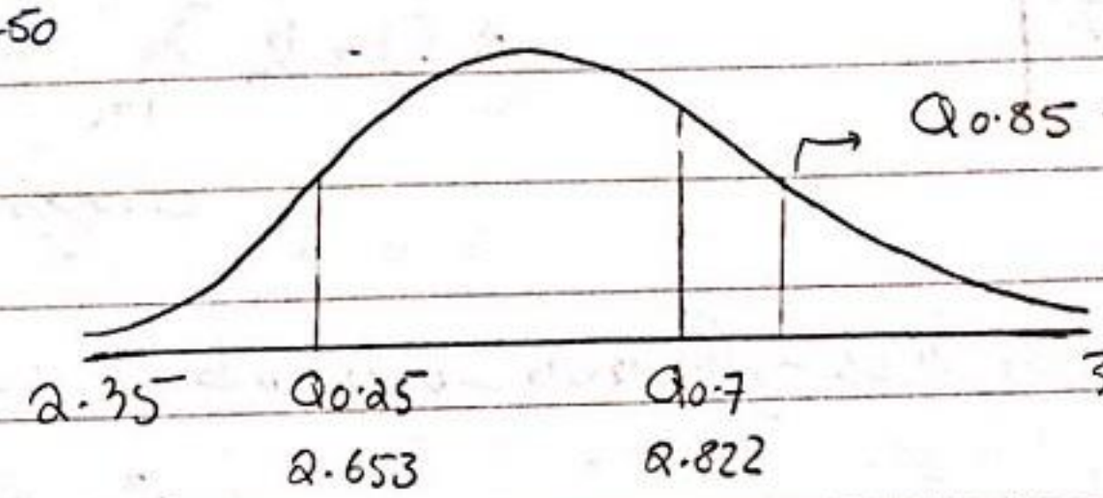
مثال: در جدول توزیع فراوانی ذیل مطلوبست محاسبه چارک اول، دهک هفتم و صدک چهل و پنجم.

	f_i	f_{ci}
2.35-2.45	1	1
2.45-2.55	4	5
2.55-2.65	7	12
2.65-2.75	15	27 → طبقه چارک اول
2.75-2.85	11	38 → طبقه دهک هفتم
2.85-2.95	10	48 → طبقه صدک چهل و پنجم
2.95-3.05	2	50
	$n=50$	

$$Q_{0.25}\begin{cases} n\cdot p = 50\times 0.25 = 12.5 \\ Q_{0.25} = 2.65 + \left(\frac{12.5-12}{15}\right)\times 0.1 = 2.653 \end{cases}$$

$$Q_{0.7}\begin{cases} n\cdot p = 50\times 0.7 = 35 \\ Q_{0.7} = 2.75 + \left(\frac{35-27}{11}\right)\times 0.1 = 2.822 \end{cases}$$

$$Q_{0.85}\begin{cases} n\cdot p = 50\times 0.85 = 42.5 \\ Q_{0.85} = 2.85 + \left(\frac{42.5-38}{10}\right)\times 0.1 = 2.895 \end{cases}$$



شاخص های پراکندگی:

یک مسئله مهم در مورد داده های آماری، سنجش میزان تغییرات و پراکندگی آن ها می باشد که به وسیله شاخص های پراکندگی توضیح داده می شوند. مثلاً فرض کنید طول عمر باتری های ساخته شده توسط ۲ کارخانه بصورت ذیل می باشد:

کارخانه اول: 100، 102، 103، 105، 107، 107، 111

کارخانه دوم: 96، 100، 104، 105، 107، 107، 116

محاسبه شاخص های تمایل به مرکز در این 2 سری نمونه داده نتایج ذیل را ایجاد می آورید

کارخانه اول: $\bar{x}_1 = 105$ $Md_1 = 105$ $MO_1 = 107$

کارخانه دوم: $\bar{x}_2 = 105$ $Md_2 = 105$ $MO_2 = 107$

از نظر شاخص های تمایل به مرکز، تفاوتی بین این 2 سری نمونه داده مشاهده نمی گردد ولی با دقت کار ملاحظه گردد می توان دریافت که $R_1 = 111-100 = 10$ و $R_2 = 116-96 = 20$. بنابراین دامنه داده های طول عمر باتری در نمونه دوم بیشتر از نمونه اول می باشد. بنابراین پراکندگی داده ها در نمونه دوم بیشتر از نمونه اول می باشد لذا محصولات کارخانه اول همگن تر یا بهتر از محصولات کارخانه دوم می باشد.

phoenix
Scanned with CamScanner

Subject:

date:

بهترین شاخص‌های پراکندگی:

۱- دامنه تغییرات (R):

$R = X_{Max} - X_{Min}$

R داده‌های اول $= 111 - 100 = 10$

R داده‌های دوم $= 116 - 96 = 20$

۲- انحراف چارک‌ها یا نیم دامنه چارک‌ها:

$$Q = \frac{1}{2}(Q_{0.75} - Q_{0.25})$$

میزان پراکندگی داده‌ها را حول میانگین نشان می‌دهد.

۳- میانگین قدرمطلق انحرافات یا انحراف متوسط: متوسط قدرمطلق انحرافات تک‌تک داده‌ها از میانگین را نشان می‌دهد.

$$A.D = \frac{1}{n}\sum_{i=1}^{n} |X_i - \bar{X}|$$

در مجموعه داده‌ها

$$A.D = \frac{1}{n}\sum_{i=1}^{K} f_i |X_{ci} - \bar{X}|$$

در جدول توزیع فراوانی

۴- واریانس و انحراف استاندارد یا متوسط مربع انحرافات داده‌ها از میانگین را واریانس می‌نامند، جذر واریانس انحراف استاندارد نامیده می‌شود.

واریانس نمونه:
$$\begin{cases} S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \\ S^2 = \frac{1}{n-1}\sum_{i=1}^{k}(X_{ci} - \bar{X})^2 \end{cases}$$

واریانس جامعه:
$$\begin{cases} \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 \\ \sigma^2 = \frac{1}{n}\sum_{i=1}^{k}(X_{ci} - \mu)^2 \end{cases}$$

انحراف استاندارد نمونه $S = \sqrt{S^2}$ (فرمول‌های ترفندی)

انحراف استاندارد جامعه $\sigma = \sqrt{\sigma^2}$

فرمول‌های کاربردی:

$$S^2 = \frac{1}{n-1}\left(\sum f_i x_{ci}^2 - \frac{(\sum f_i x_{ci})^2}{n}\right)$$

$$\sigma^2 = \frac{1}{n}\left(\sum f_i x_{ci}^2 - \frac{\sum (f_i x_{ci})^2}{n}\right)$$

۵- ضریب تغییرات (C.V): برای مقایسه پراکندگی داده‌ها با یکی از شاخص‌های استفاده شده جهت اندازه‌گیری داده‌ها بستگی ندارد باشد. یکی از این شاخص‌ها، ضریب تغییرات (C.V) می‌باشد که به صورت درصدی بیان می‌شود:

$$C.V = \frac{\sigma}{\mu} \times 100$$

$$C.V = \frac{S}{\bar{X}} \times 100$$

نمونه

Scanned with CamScanner

Subject:

date:

مثال: جدول زیل سمپل با طول قد 100 دانشجوی پسر است مطلوبست محاسبه:

الف) اندازه متوسط ب) اندازه چارک ها ج) واریانس و انحراف معیار د) ضریب تغییرات

روش حل بر اساس روش کدگذاری با تغییر متغیر

$u_i = \frac{x_{ci} - A}{c}$ ← مرکز طبقه i ام ، A ← مرکز طبقه وسط ، c ← طول طبقات

	f_i	x_{ci}	u_i	$f_i u_i$	$\lvert u_i - \bar{u} \rvert$	$f_i \lvert u_i - \bar{u} \rvert$	$f_i u_i^2$	f_{ci}
149.5-156.5	15	153	-2	-30	+1.95	29.25	60	15
156.5-163.5	20	160	-1	-20	+0.95	19	20	35
163.5-170.5	30	167	0	0	0.05	1.5	0	65
170.5-177.5	25	174	1	25	1.05	20.25	25	90
177.5-184.5	10	181	2	20	2.05	20.5	40	100
	100			-5		96.5	145	

$\bar{u} = \frac{1}{n}\sum_{i=1}^{5} f_i u_i = -5/100 = -0.05$ $\Rightarrow$ $\bar{x} = A + C\bar{u} = 167 + 7(-0.05) = 166.65$

$A.D_u = \frac{1}{n}\sum f_i \lvert u_i - \bar{u} \rvert = \frac{96.5}{100} = 0.965$ $\Rightarrow$ $A.D_x = C \times A.D_u = 7 \times 0.965 = 6.755$

$Q = \frac{1}{2}(Q_{0.75} - Q_{0.25})$ $\Rightarrow$ $Q_{0.25} = \begin{cases} 100 \times 0.25 = 25 \\ Q_{0.25} = 156.5 + \left(\frac{25-15}{20}\right) \times 7 = 160 \end{cases}$

$Q_{0.75} = \begin{cases} n.P = 100 \times 0.75 = 75 \\ Q_{0.75} = 170.5 + \left(\frac{75-65}{25}\right) \times 7 = 173.3 \end{cases}$ $\Rightarrow$ $Q = \frac{1}{2}(173.3 - 160) = 6.65$

$\delta_u^2 = \frac{1}{n-1}\left(\sum f_i u_i^2 - \frac{(\sum f_i u_i)^2}{n}\right) = \frac{1}{99}\left(145 - \frac{(-5)^2}{100}\right) = 1.462$

$S_x^2 = C^2 \times S_u^2 = (7^2) \times 1.462 = 71.638$ $\Rightarrow$ $S_x = 8.4639$

$C.V = S/\bar{x} \times 100 = \frac{8.4639}{166.65} \times 100 = 5.0788$

Scanned with CamScanner

Subject:

date:

مثال: با توجه به جدول توزیع فراوانی ذیل مطلوبست محاسبه:

- میانگین
- میانه
- واریانس
- ضریب تغییرات
- دهک هفتم

حدود طبقات	10-20	20-30	30-40	40-50
f_i	1	3	4	2

	x_{ci}	f_i	f_{ci}	$f_i x_{ci}$	$f_i x_{ci}^2$
10-20	15	1	1	15	225
20-30	25	3	4	75	1875
30-40	35	4	8	140	4900
40-50	45	2	10	90	4050
				320	11075

$$\bar{X} = 1/n \sum f_i x_{ci} = 320/10 = 32$$

$$Md = 30 + \left(\frac{5-4}{4}\right) \times 10 = 32.5$$

$$S^2 = 1/9 \left(11075 - \frac{(320)^2}{10} \right) = 1/9 \ (11075 - 10240) = 92.78$$

$$S = 3.120$$

$$C.V = \frac{3.120}{32} \times 100 = \%10.03$$

$$Q_{0.7} = \begin{cases} n \cdot p = 10 \times 0.7 = 7 \\ Q_{07} = 30 + \left(\frac{7-4}{4}\right) \times 10 = 37.5 \end{cases}$$

مثال: جدول ذیل مربوط به آب مصرفی 20 خانوار است، مطلوبست محاسبه:

الف) انحراف چارک ها
ب) انحراف از میانگین
ج) واریانس و انحراف از معیار
د) ضریب تغییرات

	f_i
9.5-12.5	2
12.5-15.5	4
15.5-18.5	6
18.5-21.5	3
21.5-24.5	5

Scanned with CamScanner