
بسم الله الرحمن الرحيم

درس آمارو احتمالات
رشته مهندسی علوم و صنایع غذایی

برگرفته از کتاب آمار کاربردی و طرح آزمایش ها برای علوم کشاورزی جلد اول تالیف دکتر سید احمد سادات نوری انتشارات دانشگاه تهران

مدرس: خانم مهندس ثمانه لواسانی

مقدمه

امروزه در علوم پایه و مهندسی به دفعات نیاز به جمع آوری اطلاعات در مورد مجموعه‌های از اشیاء و یا انسانها داریم که این امر بر عهده علم آمار می‌باشد و با کمک آمار توصیفی می‌توان اطلاعات جمع آوری شده را به صورتی منظم گرد آوری نمود بطوریکه بتوان با یک نگاه اجمالی به نتایج بدست آمده یک دید کلی نسبت به کل داده‌ها بدست آورد.

تعاریف و مفاهیم پایه

۱-۱ تعاریف اولیه

جامعه آماری : به مجموعه‌ای از اشیاء یا افراد که حداقل یک ویژگی مشترک آنها مورد مطالعه قرار می‌گیرد. جامعه آماری می‌گوییم. ویژگی‌های مشترک یک جامعه آماری از عضوی به عضو دیگر تغییر می‌کند. که آنها را متغیر می‌نامند. متغیرها معمولاً به دو نوع تقسیم می‌شوند:

متغیر کمی : متغیرهایی می‌باشند که معمولاً قابل اندازه‌گیری هستند و می‌توان مقدار آنها را به صورت عددی نمایش داد مثل مقدار وزن، قد، حجم و ....

۱- متغیر کیفی : متغیرهایی می‌باشند که مستقیماً توسط اعداد و ارقام قابل اندازه‌گیری نیستند. مثل گروه خونی، شغل، رنگ چشم و ... که برای اندازه‌گیری این متغیرها به آنها عددی نسبت می‌دهیم.

متغیرهای کمی خود بر دو نوع هستند.

۱- گسسته : متغیرهایی که بین دو مقدار متصور آنها هیچ عدد دیگری وجود نداشته باشد.

۲- پیوسته : متغیرهایی که بین هر دو مقدار متصور آنها همواره عددی دیگری وجود دارد. مثل وزن یا طول و قد افراد.

پس از جمع آوری داده‌ها برای رسیدن به اهداف مورد نیاز به بررسی و تجزیه و تحلیل داده‌ها داریم که برای این منظور ابتدا داده‌ها را در یک جدول تنظیم و طبقه‌بندی می‌کنیم و سپس با استفاده از نمودارهای آماری نحوه توزیع داده‌ها را نمایش می‌دهیم و در نهایت داده‌ها را با کمک چند عدد به نام شاخص یا آماره خلاصه می‌کنیم.

انواع مقیاس ها

۱- مقیاس اسمی: این مقیاس برای شناسایی افراد یا اشیا یا مکانها به کار می رود. مثال: اگر دانشجویان دانشگاهی از شهرهای مختلف مثل تهران، شیراز، اصفهان، مازندران باشند به ترتیب با اعداد ۱،۲،۳،۴ مشخص میشوند.

۲- مقیاس ترتیبی: این مقیاس بر اساس مهارت طبقه بندی میشوند مثل مهارت ۴ کارگر کارخانه ای که با اعداد ۱تا ۴ مشخص میشود.

۳- مقیاس نسبتی: هرگاه مقیاس سنجش نسبت را حفظ کند مقیاس نسبتی گویند.

نمادهای مهم درآمار در جدول ارائه شده

شاخص مورد مطالعه	نمونه (آماره)	جامعه (پارامتر)
میانگین	$\overline{X}$	μ (مو)
نسبت	ρ	$\hat{P}$ (پی هد)
همبستگی	r_{xy}	ρ_{xy} (رو)
واریانس	S^2	δ^2 (سیگما به توان دو)
انحراف معیار	S	δ (سیگما)
تعداد	N	N

X = نمره یا نمره ها

L = کرانه طبقات

F_c = فراوانی تجمعی

X^2 = مجذور(مربع نمره) انحراف نمره از میانگین

MS = میانگین مجموع مجذور انحراف نمره از میانگین

X_c = حد میانی نمرات

F_i = فراوانی مطلق

x = انحراف نمره از میانگین

SS = مجموع مجذورات انحراف نمره از میانگین

i = فاصله طبقاتی

علامت جمع در آمار بصورت $\sum$ می باشد و به صورت زیر تعریف میشود.

$$\sum_{i=1}^{N} Xi = x_1 + x_2 + \cdots + x_N$$

انواع فراوانی

برای خلاصه و جمع و جور کردن آمار خام از جدول توزیع فراوانی استفاده می شود.

۱- فراوانی مطلق: عبارتست از فراوانی x_i از متغیر x که فراوانی ساده نیز نامیده میشود.

۲- فراوانی نسبی: از تقسیم فراوانی مطلق تک تک کلاس ها بر تعداد کل داده ها یا جمع کل فراوانی مطلق بدست می اید.

$$\text{فراوانی نسبی} = \frac{\text{فراوانی مطلق}}{\text{جمع کل فراوانی}}$$

هم چنین اگر فراوانی نسبی را در عدد ۱۰۰ ضرب کنید درصد وقوع هر داده را در میان کل داده‌ها بدست می‌آورید.

۳- فراوانی تراکمی یا تجمعی:

توزیع فراوانی تراکمی: اگر پژوهشگری علاقمند به دانستن تعداد افراد یا نمره هایی باشد که در پایین نمره یا عدد خاصی وجود دارند، نیاز به توزیع فراوانی تراکمی دارد. فراوانی تراکمی با (cf) نشان داده می شود که از جمع‌کردن فراوانی‌های ساده هر طبقه با طبقه بزرگتر به دست می آید.

نکته۱: فراوانی تراکمی کوچکترین طبقه همیشه برابر با فراوانی ساده یا مطلق آن طبقه است.

نکته۲: فراوانی تراکمی بزرگترین طبقه همیشه برابر با مجموع داده ها (ΣF) یا N می باشد.

درصد فراوانی مطلق و تراکمی

$$P = \frac{\text{فراوانی مطلق هر طبقه}}{\text{تعداد کل فراوانی ها}} \times ۱۰۰ \qquad P = \frac{F}{N} \times ۱۰۰$$

$$Cf\% = \frac{\text{فراوانی تراکمی هر طبقه}}{\text{تعداد کل فراوانی ها}} \times ۱۰۰ \qquad CF\% = \frac{CF}{N} \times ۱۰۰$$

نمایده طبقات (نقاط وسط طبقات): نماینده طبقات یا نقاط میانی را با x' نمایش می دهند و ازطریق فرمول زیر به‌دست می‌آید:

$$X' = \frac{\text{حد بالای طبقه+حد پایین طبقه}}{۲}$$

(جدول توزیع فراوانی سطح هوشی ۱۰۰ دانشجو)مثال

طبقات	(فراوانی مطلق)F	cf (فراوانی تراکمی نزولی)	(فراوانی نسبی)P_f	(درصد فراوانی نسبی)P	(فراوانی تراکمی درصدی)P_{cf}	(نماینده کلاس)X_c	حدود واقعی طبقات
134-130	3	105	0.03	3	100	132	129.5-134.5
129-125	3	102	0.03	3	97.1	127	129.5-124.5
124-120	3	99	0.03	3	94.3	122	124.5-119.5
119-115	5	96	0.05	5	91.4	117	119.5-114.5
114-110	12	91	0.11	11	86.6	112	114.5-109.5
109-105	11	79	0.10	10	75.2	107	109.5-104.5
104-100	15	68	0.14	14	65	102	104.5-99.5
99-95	10	53	0.09	9	50.5	97	99.5-94.5
94-90	11	43	0.10	10	41	92	94.5-89.5
89-85	9	32	0.08	8	30.5	87	89.5-84.5
84-80	7	23	0.07	7	22	82	84.5-79.5
79-75	4	16	0.04	4	15.2	77	79.5-74.5
74-70	6	12	0.06	6	9.5	72	74.5-69.5
69-65	4	6	0.04	4	5.7	67	69.5-64.5
64-60	2	2	0.02	2	2	62	64.5-59.5
		N=105	1	100			

انواع نمودارهای اماری

نمودار دایره‌ای: نموداری است که با داده های اسمی و کیفی بکار می رود. برای محاسبه درجه و زاویه مرکزی متعلق به یک گروه از فرمول زیر استفاده می کنیم:

$$\text{درجه}^{\circ} = \frac{n}{N} \times ۳۶۰$$

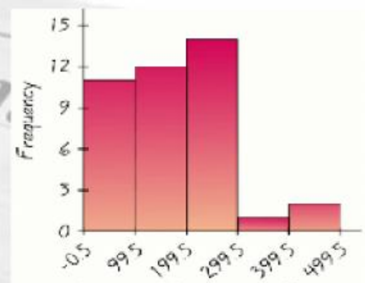
مثال: اگر داخل یک گروه ۱۲۰ نفری۶۰ نفر قبول، ۴۰ نفر تجدید و۲۰ نفر مردود شده باشند، درجه مربوط به تجدیدی‌ها چند است؟

$$\text{درجه تجدیدی} = \frac{۴۰}{۱۲۰} \times ۳۶۰ = ۱۲۰$$

نمودار ستونی (میله‌ای): نموداری است که با داده های اسمی بکار می رود که در محور عمودی، فراوانی و در محور افقی طبقات قرار می گیرند. در نمودار میله ای فاصله بین اعداد و نقاط یکسان و ثابت است. مثل تعداد فرزندان. ولی اگر برای متغیرهای طبقه ای مانند قبولی– مردودی، رنگ چشم و اعتقادات مذهبی نمودار رسم شود، هیچ ضرورتی ندارد که فاصله ها در محور X ثابت می‌ماند. برای این متغیرها از نمودار دایره ای و مانند آن استفاده می کنند (نمودار پای). مثال:

نمودار هیستوگرام: با داده های فاصله ای و نسبتی به کار می رود. نموداری شبیه نمودار ستونی است، ولی در آن ستونها به هم چسبیده است و در محور افقی (X) کرانه (حدود واقعی) طبقات و در محور عمودی (y) فراوانی مطلق قرار می گیرد.

	i
۴۰۰–۴۹۹	۲
۳۰۰–۳۹۹	۱
۲۰۰–۲۹۹	۱۴
۱۰۰–۱۹۹	۱۲
۰–۹۹	۱۱



- منظور از معیار تمایل به مرکزی یا اندازگیری گرایش مرکزی یک گروه داده، عددی است که توسط آن عدد، مرکزیت آن گروه داده مشخص می گردد.

- مهمترین معیارهای تمایل مرکزی عبارت است از :
  - میانگین حسابی
  - میانه
  - نما

۱-۴-۱ پارامترهای مرکزی

عموماً داده‌ها در جامعه آماری یکنوع تجمع .و فشردگی حول یک مقدار خاص از صفت مورد مطالعه را بوجود می‌آورند که این مقدار خاص به عنوان یک پارامتر مرکزی معرفی می‌شود. مهمترین پارامترهای مرکزی عبارتند از: میانگین، میانه و مد (نما)

شاخص های گرایش مرکزی

نما (مد)

بی ثبات ترین شاخص گرایش مرکزی است که با داده های اسمی بکار می رود و عددی است که دارای بیشترین فراوانی می باشد. مثلاً: ۹-۵-٦ -٦-٦-٦-٦-٦-٦-۷-٦-۵-۷-٦-۵-۱

در اینجا نما عدد ٦ است. برخی موارد توزیع، دونمایی یا چندنمایی می شود مثل: ۷-٦-۵-۵-٤-۲-۳-۲-۱

در اینجا ۲ و ۵ نما هستند که نما در اصل $\frac{۲+۵}{۲} = ۳/۵$ است.

در جدول فراوانی، نما عبارتست از عددی که F_i آن بزرگتر است.

در اینجا عدد ٤ نما است، زیرا عدد چهار، ۹ بار تکرار شده است.

X	F_i
۵	۵
۴	۹
۳	۳
۲	۲
۱	۶

نکته ۱: برای محاسبه نما باید ابتدا اعداد را از کوچک به بزرگ مرتب کنیم.

حدود واقعی: حدود واقعی نمرات بصورت کم کردن ۰/۵ نمره در اعداد صحیح و در اعداد اعشاری، نیم واحد یعنی ۰/۰۵– ۰/۰۰۵ و مانند آن کسر می شود. یعنی: کرانه عدد ۲۵ ←۲٤/۵ تا ۲۵/۵ و کرانه عدد ۲۵/۵ ←۲۵/٤۵ تا ۲۵/۵۵ می‌باشد. مفهوم حدود واقعی مخصوصاً زمانی مفید است که اعداد گروه بندی یا طبقه بندی شوند. مثال: پس از اجرای یک آزمون ریاضی مشاهده می شود که ۱۰ نفر نمره ۱۲ گرفته اند. این بدان معنی نیست که همه توانایی یکسان دارند، بلکه دقیق نبودن وسیله اندازه گیری ممکن است موجب این امر شده باشد. به این خاطر نیاز به حدود واقعی می‌باشد؛ یعنی ۱۲/۵– ۱۱/۵.

محاسبه نما در داده ای طبقه بندی شده

از معادله‌ی زیر محاسبه می شود، که در آن:

L = حد واقعی پایین طبقه ای که دارای بیشترین فراوانی است

i = طول یا فاصله طبقات $R = X_H - X_L + ۱$

$d_۱$ = تفاوت فراوانی طبقه نمادار از فراوانی طبقه قبل

$d_۲$ = تفاوت فراوانی طبقه نمادار از طبقه بعد.

مثال:

X	F
۵۴-۵۶	۱
۵۷-۵۹	۳
۶۰-۶۲	۶
۶۳-۶۵	۸
۶۶-۶۸	۲

طبقه نما ←

$$MO = L + i\left(\frac{d_۱}{d_۱ + d_۲}\right)$$

$$d_۱ = ۸ - ۶ = ۲$$

$$d_۲ = ۸ - ۲ = ۶$$

$$MO = ۶۲/۵ + ۳\left(\frac{۲}{۲ + ۶}\right) = ۶۳/۲۵$$

نکته: چنانچه توزیع نرمال باشد، نما از فرمول زیر محاسبه می شود: $MO = ۳Md - ۲\overline{X}$

میانه

میانه جایگاهی در توزیع نمره هاست و توزیع نمره ها را به دو قسمت مساوی تقسیم می کند؛ یعنی جایی است که دقیقاً ۵۰ درصد نمره ها بالای آن و ۵۰ درصد نمره ها زیر آن قرار می گیرند. میانه از نما باثبات‌تر و از میانگین بی‌ثبات‌تر است و با داده های رتبه ای بکار می رود. زیرا ما ابتدا نمره ها را از کوچک به بزرگ مرتب می کنیم.

طریقه محاسبه میانه نمرات خام در اعداد گسسته

ابتدا نمره ها را از کوچک به بزرگ مرتب می کنیم و سپس اگر تعداد اعداد فرد باشد، میانه، عدد وسط است.

مثال: ۱۲-۹-٤-۳-۱-٦-٥

مثال: ۱۲-۹-٦-٥-٤-۳-۱ در اینجا عدد ٥ میانه است

مثال: در توزیع اعداد ۳۱-٤-٥-۲۱-۳۰-۳۰-۱۷-۸-۱۲ میانه برابر است با ۱۷.

طریقه محاسبه میانه اعداد خام در تعداد زوج

۱- اعداد را از کوچک به بزرگ مرتب می کنیم.

۲- دو عدد وسط را با هم جمع و تقسیم بر دو می کنیم. ۱۲-۹-۲-۸-٥-۳

٥+۸ برابر ۱۳ و ۱۳ تقسیم بر ۲ میانه محاسبه می شود که برابر ٦/٥ می باشد. ۱۲-۹-۸-٥-۳-۲

محاسبه میانه در جدول اعداد طبقه بندی شده

۱- ابتدا نمرات را از کوچک به بزرگ مرتب می کنیم.

۲- سپس توزیع فراوانی و جدول را تشکیل و با فرمول نمرات طبقه بندی شده میانه را محاسبه می کنیم.

مثال: ۳-۵-۲-۴-۸-۴-۴-۴-۴-۲

X	F_i	F_c
۸	۱	۱۰
۵	۱	۹
۴	۵	۸
۳	۱	۳
۲	۲	۲

L= حد پایین طبقه میانه دار $md = L + \dfrac{\frac{N}{۲} - F_C}{F_i} i$

طبقه‌ی میانه دار= مجموع فراوانی تقسیم بر دو= $\dfrac{N}{۲}$

F_c = فراوانی تجمعی ماقبل سطری که در آن میانه واقع شده است.

F_i = فراوانی مطلق طبقه میانه دار

i = فاصله طبقاتی

مثال:

F_c=۳ , F_i=۵ , i=۱

$$md = L + \frac{\frac{N}{۲} - F_c}{F_i} \times i \quad \Rightarrow \quad md = ۳/۵ + \frac{۵ - ۳}{۵} \times ۱ \qquad md = ۳/۹$$

میانگین

میانگین، باثبات‌ترین شاخص گرایش مرکزی است که با داده‌های فاصله‌ای و نسبی بکار می‌رود و مرکز ثقل داده‌هاست.

طریقه محاسبه میانگین اعداد خام

مثال: ۱۱-۹-۱۰

$$\overline{X}=\frac{\sum X}{N} \Longrightarrow \text{میانگین} = \frac{\text{مجموع نمره ها}}{\text{تعداد نمره ها}}$$

$$\overline{X}=\frac{۱۰+۹+۱۱}{۱}=۱۰ \qquad \overline{X}=۱۰$$

طریقه محاسبه اعداد طبقه بندی شده با i=۱

$$\bar{X}=\frac{\sum F.X}{n}$$

مثال:

X	Fi
۵	۳
٤	۸
۳	۵
۲	۱۲
	$\sum F_i = n = ۲۸$

F.X
۱۵
۳۲
۱۵
۲٤
$\sum F.X = ۸٦$

$$\overline{X}=\frac{۸٦}{۲۸}=۳/۰۷$$

۱-۱۲. مثال: فرض کنید دانشجویی نمرات زیر را در دروس ترم قبل خویش گرفته باشد.

نام درس	تعداد واحد	نمره
آمار و کاربرد آن	۴	۱۹
جغرافیای اقتصادی	۲	۱۶
اقتصاد کلان ۱	۳	۱۸
اقتصاد خرد ۲	۴	۱۲
اقتصاد کشاورزی	۳	۱۴

در این مثال، تعداد واحدها، وزن‌های مربوطه را نشان می‌دهد. بنابراین، میانگین وزنی را به صورت زیر می‌توان محاسبه کرد.

$$\mu_w = \frac{(۴\times ۱۹)+(۲\times ۱۶)+(۳\times ۱۸)\times(۴\times ۱۲)+(۳\times ۱۴)}{۱۶} = \frac{۲۵۲}{۱۶} = ۱۵/۷۵$$

طریقه محاسبه میانگین اعداد طبقه بندی شده با ۱ ≠i

$$\overline{X} = \frac{\sum F.X_c}{n}$$

مثال:

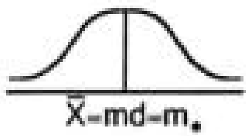
X	Fi		Xc	F.X
۱۷–۱۹	۵		۱۸	۹۰
۱۴–۱۶	۶	$\Longrightarrow$	۱۵	۹۰
۱۱–۱۳	۱۱		۱۲	۱۳۲
۸–۱۰	۳		۹	۲۷
۵–۷	۹		۶	۵۴
	$\sum F_i =$ ۳۴			$\sum F.X_c =$ ۳۹۳

$$\overline{X} = \frac{۳۹۳}{۳۴} = ۱۱/۵۶ \qquad \overline{X} = ۱۱/۵۶$$

نکته

۱- اگر میانگین و میانه و نما با هم برابر باشد، توزیع متقارن است و کجی آن صفر است.

$\overline{X} = md = mo$

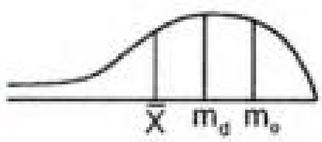


۲- اگر میانگین بزرگتر از میانه و میانه بزرگتر از نما باشد کجی مثبت است، یعنی اکثر نمرات پایین بوده و امتحان سخت است.

$\overline{X} > md > m_o$

۳- اگر میانگین کوچکتر از میانه و میانه کوچکتر از نما باشد، کجی منفی است و اکثر نمرات بالاست و امتحان آسان بوده است.

$\overline{X} < md < mo$



میانگین هارمونیک (همساز)

این نوع میانگین برای مواردی بکار می رود که مقیاس ترکیبی باشد مانند متر در ثانیه و کیلومتر بر ساعت. این میانگین در عینک‌سازی و مطالعه شبکه های برقی به کار می رود.

$$\mathrm{HM} = \frac{۱}{\frac{۱}{N}\left[\frac{۱}{x_۱}+\frac{۱}{x_۲}+\ldots+\frac{۱}{x_n}\right]}$$

مثال: میانگین سرعت های ۵ و۶ و۷ و۲ کیلومتر در ساعت ٤ ماشین چند است؟

$$\mathrm{HM} = \frac{۱}{\frac{۱}{٤}\times\left[\frac{۱}{۵}+\frac{۱}{۶}+\frac{۱}{۷}+\frac{۱}{۲}\right]} = ۳/۹۶$$

۱–۱۸. مثال: جدول توزیع فراوانی مصرف برق ۵۰ خانوار به صورت زیر داده شده است، میانگ هارمونیک را حساب کنید.

x_i	f_i	حدود طبقات
۵۱	۴	۵۰–۵۲
۵۴	۹	۵۳–۵۵
۵۷	۲۱	۵۶–۵۸
۶۰	۱۴	۵۹–۶۱
۶۳	۲	۶۲–۶۴

$$H = \frac{n}{\sum_{i=۱}^{n}\frac{f_i}{x_i}} = \frac{۵۰}{\left(\frac{۴}{۵۱}+\frac{۹}{۵۴}+\frac{۲۱}{۵۷}+\frac{۱۴}{۶۰}+\frac{۲}{۶۳}\right)} \cong ۵۶٫۹۵$$

میانگین هندسی

نوعی میانگین است که با G نشان داده می شود و معمولاً هرگاه x_i ها از درصدها یا نسبت ها تشکیل شده باشند، استفاده می‌شود. در کارهای اقتصادی یا جمعیت شناسی به کار می رود.

$$G=\sqrt[n]{X_1.X_2.X_3(X_n)}$$

مثال: میزان سود شرکت مهرداد در ۵ سال گذشته برحسب درصد به ترتیب ۲، ۳، ٤، ٤، ۳، شده است. کدام‌یک از گزینه های زیر وضع سود آوری شرکت را بهتر نشان می دهد.

$$G=\sqrt[5]{3\times 4\times 4\times 2\times 3}=3/01$$

مثال ۲: میزان خسارت در محصولی در چهار مرحله نمونه برداری بصورت ۳۰، ۲۵، ۲۴، ۴۵٪ می باشد میانگین هندسی را محاسبه نمایید؟

$$\sqrt[4]{45\times 24\times 25\times 30}=30\%$$

رابطه بین سه میانگین فوق به‌صورت زیر است:

$$\overline{X}>G>HM$$

میانگین هارمونیک > میانگین هندسی > میانگین وزنی

حال اگر تعداد داده ها به علاوه ۱ را بر ۴ تقسیم کنیم چارکها بدست میآیند.
این چارک ها برای رسم نمودار جعبه ایی ضروری هستند. چارک ها داده ها را به چهار بخش مساوی تقسیم میکنند.

چارکها...

- چارک اول $$\frac{n+1}{4}$$
- چارک دوم یا میانه $$\frac{2(n+1)}{4}$$
- چارک سوم $$\frac{3(n+1)}{4}$$

چندک ها

- اگر مخرج فرمول قبل را 10 قرار دهیم دهک ها، اگر 100 قرار دهیم صدک ها بدست میایند.

- مثال: دهک 3 و 5 و 9 اینگونه بدست میایند.

$$\frac{3(n+1)}{10} \qquad \frac{5(n+1)}{10} \qquad \frac{9(n+1)}{10}$$

- مثال: صدک 30 و 50 و 75 اینگونه بدست میایند.

$$\frac{30(n+1)}{100} \qquad \frac{50(n+1)}{100} \qquad \frac{75(n+1)}{100}$$

پراکندگی یکی از مهم‌ترین مفاهیم در آمار است. هدف از اندازه‌گیری معمولاً پیدا کردن تغییرات و توجیه آن‌هاست. هرچه پراکندگی کمتر باشد، پیش‌بینی مقدار یک متغیر تصادفی با کمک مقدار میانگینش دقیق‌تر می‌شود؛ به عبارت دیگر، پراکندگی می‌تواند دقتِ یک پیش‌بینی را نشان دهد.

شاخص های پراکندگی

هروقت یک شاخص مرکزی به ما دادند برای اینکه ببینیم این شاخص مرکزی چقدر دلچسب است شاخص پراکندگی را در نظر می گیریم تا ببینیم چقدر به این شاخص اعتماد داشته باشیم

- دامنه
- واریانس
- انحراف استاندارد
- ضریب تغییرات

دامنه (RANGE)

- تفاوت بین بزرگترین و کوچکترین داده.
- به میزان زیادی تحت تأثیر داده های پرت قرار می‌گیرد.
- برای داده‌های متقارن بدون هیچ برون‌هشتی مناسب است.

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

یکی از روش‌های اندازه‌گیری پراکندگی بین داده‌ها، محاسبه «دامنه تغییرات» است. این شاخص، حداکثر میزان پراکندگی را نشان می‌دهد و برای محاسبه آن کافی است که تفاوت بین بزرگترین و کوچکترین مقدار را بدست آورد.

مثال: در توزیع نمرات ۳-۲۰-۱۴-۱۵-۹-۶-۵ دامنه تغییرات را محاسبه کنید:

R= ۲۰-۳+۱=۱۸ R= ۱۸

دامنه میان چارکی

برای آنکه بتوان مشکل تاثیر پذیری دامنه تغییرات از مقدارهای بزرگ و کوچک را از بین برد، می‌توان فاصله بین بزرگترین و کوچکترین مقدار را براساس چارک‌ها محاسبه کرد. به این ترتیب برای داده‌هایی که دارای مقدارهای دور افتاده هستند، فاصله بین چارک اول و سوم، می‌تواند برآورد بهتری برای محاسبه حداکثر پراکندگی داده‌ها بدست دهد. به این شاخص «دامنه میان چارکی» می‌گویند. شکل محاسباتی دامنه میان چارکی به صورت زیر است:

IQR=Q3−Q1

که در آن Q_1 چارک اول و Q_3 چارک سوم است.

مثال: برای نمرات ۶۰۰–۳۲–۱۵–۹–۶–۵–۳ انحراف چارکی مناسب است.

$Q_۱$ (۵) md (۹) $Q_۳$ (۳۲)

$$SIRQ = \frac{Q_۳ - Q_۱}{۲} = \frac{\text{چارک اول – چارک سوم}}{۲}$$

$$Q = \frac{۳۲ - ۵}{۲} = ۱۳/۵$$

متوسط قدر مطلق انحرافات

انحراف متوسط (MD)

انحراف متوسط، یک شاخص پراکندگی است که به آن میانگین قدر مطلق انحراف نمره از میانگین گفته می‌شود

$$\mathrm{MD} = \frac{\sum\left|X-\overline{X}\right|}{N}$$

مثال: برای اعداد ۵–۴–۳–۲–۱ خواهیم داشت:

X	$X-\overline{X}$	$\left\|X-\overline{X}\right\|$
۵	۲	۲
۴	۱	۱
۳	۰	۰
۲	−۱	۱
۱	−۲	۲
		$\sum\left\|X-\overline{X}\right\|=٦$

$$md=\frac{٦}{۵}=۱/۲$$

واریانس (VARIANCE)

- منظور از واریانس ،بررسی پراکندگی داده های کمی در اطراف میانگین داده ها است .
- این پراکندگی ها هر چقدر در اطراف و نزدیک میانگین قرار بگیرند ، شرایط مطلوب تر خواهد بود.

$$\delta^2 = \frac{\sum (x-\bar{x})^2}{n}$$

انحراف استاندارد(SD)

- انحراف معیار نمونه ریشهٔ دوم واریانس نمونه می‌باشد.
- واحدها، واحدهای اصلی هستند
- انحراف متوسط داده‌ها از میانگین خود را اندازه می‌گیرد.
- همچنین به میزان زیادی تحت‌تأثیر داده های پرت قرار دارد.

$$\delta=\sqrt{\delta^2}$$

نکته: توان دوم انحراف معیار، واریانس نام دارد. انحراف معیار، باثبات‌ترین شاخص پراکندگی است. هرچقدر انحراف استاندارد بیشتر باشد، پراکندگی بیشتر است. شاید اساسی ترین فایده انحراف استاندارد این باشد که با استفاده از آن می توان مشخص کرد چه نسبتی از نمره ها در فاصله های مختلف نسبت به میانگین قرار گرفته است.

تصحیح شپرد

فرمول شپرد برای تصحیح انحراف معیار زمانی کاربرد دارد که فاصله طبقاتی بزرگ و تعداد طبقات کمتر از ۱۲ باشد.

$$S_c = \sqrt{S^2 - \frac{i^2}{12}}$$

مثال: اگر فاصله طبقاتی ۱۰ و انحراف معیار توزیع ۵ باشد، انحراف معیار تصحیح شده برابر خواهد بود با:

$$S_c = \sqrt{25 - \frac{10^2}{12}} = 4/97$$

ضریب تغییرات

- اگر هدف مقایسه پراکندگی توزیع وزن بدن در دو جامعه باشد که در یکی واحد اندازه گیری کیلوگرم و در دیگری پوند است، چون انحراف معیار برای جامعه اول برحسب کیلوگرم و در جامعه دوم برحسب پوند بیان خواهد شد در نتیجه این مقایسه بر اساس انحراف معیار منطقی نخواهد بود.

- به منظور رفع این اشکال از نسبت انحراف معیار به میانگین که معمولا بصورت درصد بیان می شود استفاده می گردد. کمیت حاصل را که یک مشخص کننده نسبی است با CV نشان می دهیم.

ضریب تغییرات...

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

اگر میانگین و انحراف معیار درجه حرارت بدن برای افراد یک جامعه به ترتیب اعداد ۳۶/۹ و ۰/۱۸ درجه باشد و میانگین و انحراف معیار تعداد ضربان نبض به ترتیب ۷۸ و ۹ بار در دقیقه باشد ، برای محاسبه ضریب تغییرات خواهیم داشت:

$$CV_1 = \frac{0.18}{36.9} \times 100 = 0.5$$

$$CV_2 = \frac{9}{78} \times 100 = 11.5$$

مثال ۶: یک کارخانه تولید لاستیک دو نوع محصول A و B تولید می‌کند. لاستیک نوع A دارای میانگین طول عمر ۲۰۰۰۰ کیلو متر و انحراف استاندارد ۲۰۰۰ کیلو متر می‌باشد و لاستیک نوع B دارای میانگین طول عمر ۱۸۰۰۰ کیلو متر و انحراف استاندارد ۲۰۰ کیلو متر می‌باشد، کدام نوع لاستیک برای خرید مناسب‌تر می‌باشد؟

$$X_A = ۲۰۰۰۰ \qquad \overline{X}_B = ۱۸۰۰۰ \qquad \Rightarrow \qquad CV_A = \frac{۲۰۰۰}{۲۰۰۰۰} = ۰/۱$$

$$S_A = ۲۰۰۰ \qquad S_B = ۲۰۰ \qquad \Rightarrow \qquad CV_B = \frac{۲۰۰}{۱۸۰۰۰۰} = ۰/۰۱$$

$$CV_A = ۰/۱ \times ۱۰۰ = \%۱۰$$

$$CV_A = ۰/۰۱ \times ۱۰۰ = \%۱$$

همانطور که ملاحظه می‌کنید میانگین طول عمر لاستیک دوم از لاستیک اول کمتر است ولی با توجه به اینکه ضریب تغییرات لاستیک دوم کمتر از لاستیک اول است خرید لاستیک دوم به صرفه‌تر می‌باشد.

مثال: کارخانه ای دو نوع لاستیک اتومبیل تولید می کند. برای نوع الف میانگین عمر ۱۰۰۰۰ کیلومتر، با انحراف استاندارد ۲۰۰۰ کیلومتر، و برای نوع ب میانگین عمر ۱۱۰۰۰ کیلومتر با انحراف استاندارد ۱۰۰۰ کیلومتر می باشد. کدام نوع لاستیک بهتر است؟

الف) $V = \frac{۲۰۰۰}{۱۰۰۰۰} \times ۱۰۰ = ۲۰$

ب) $V = \frac{۱۰۰۰}{۱۱۰۰۰} \times ۱۰۰ = ۹$

پاسخ: نوع (ب) بهتر است، زیرا هم میانگین عمر آن بیشتر است و هم ضریب تغییر آن کوچکتر.

نمره های استاندارد

نمره های استاندارد موقعیت فرد را در گروه معین می کنند. با داده‌های مقیاس حداقل فاصله ای کاربرد دارند. در تبدیل نمرات خام به نمره استاندارد از فرمول مقابل استفاده می کنیم:

$$Z = \frac{X - \overline{X}}{S}$$

مثال: در امتحانی که میانگین آن ۳۸ و انحراف معیار آن ۳ باشد، کسی که نمره ٤٤ گرفته است دارای ۲= Z می‌باشد.

$$Z = \frac{٤٤ - ٣٨}{٣} = \frac{٦}{٣} = ٢$$

نمره های استاندارد	میانگین	انحراف استاندارد
Z	۰	۱
T	۵۰	۱۰

- نمره Z فردی ۱/۵− شده است. نمره t او چقدر است؟

$$t = ۵۰+(-۱/۵)\times ۱۰ = ۳۵$$

- نمره t فردی ۳۰ شده است. نمره Z او چقدر است؟

$$Z = \frac{۳۰ - ۵۰}{۱۰} = -۲$$

$$t = ۵۰ + ۱۰\, z$$

۱-۱۳ استاندارد سازی

از یک جامعه آماری n نمونه $x_1, x_2, \ldots, x_n$ بصورت تصادفی انتخاب می‌کنیم بطوریکه میانگین و واریانس نمونه‌ها بترتیب $\bar{x}$ و S_x^2 می‌باشد. با توجه به تغییر مبدأ و مقیاس مقدار هر نمونه را از میانگین نمونه‌ها کم می‌کنیم و حاصل را بر S_x تقسیم می‌کنیم بنابر این داده‌های

$$y_1 = \frac{x_1 - \bar{x}}{S_x} \quad , \quad y_2 = \frac{x_2 - \bar{x}}{S_x} \quad , \quad y_n = \frac{x_n - \bar{x}}{S_x}$$

را خواهیم داشت.

با محاسبه مقایر میانگین و واریانس داده‌های جدید داریم:

$$\bar{y} = \frac{\bar{x} - \bar{x}}{S_x} = 0 \quad ; \quad S_y^2 = \left(\frac{1}{S_x^2}\right) S_x^2 = 1$$

همانطور که ملاحظه می‌کنید داده‌های جدید دارای میانگین صفر و واریانس ۱ می‌باشند که به آنها داده‌های استاندارد شده می‌گوییم. همینطور اگر داده‌های x_1 تا x_n را با متغییر تصادفی X نمایش دهیم در این صورت $Y = \dfrac{X - \bar{x}}{S_x}$ فرم استاندارد شده یا صورت معیاری متغیر تصادفی X می‌باشد.

استاندارد سازی

مثال : نمره علی از امتحان فیزیک و ریاضی به ترتیب برابر ۴۰ و ۶۰ شده است اگر میانگین نمرات امتحان فیزیک و ریاضی به ترتیب برابر ۲۰ و ۵۰ باشد و انحراف معیار امتحان فیزیک و ریاضی به ترتیب برابر ۱ و ۲ باشد علی کدام درس را بهتر امتحان داده است.

حل: برای اینکه بتوان نمرات دو درس را با یکدیگر مقایسه نمود می‌بایستی ابتدا نمرات را استاندارد سازی نمود و سپس آنها را با یکدیگر مقایسه نمود.

نمرات استاندارد شده علی در درس ریاضی $= \frac{۶۰ - ۵۰}{۲} = ۵$

نمرات استاندارد شده علی در درس فیزیک $= \frac{۴۰ - ۲۰}{۱} = ۲۰$

با وجود اینکه نمره علی در درس فیزیک کمتر از ریاضی می‌باشد اما با استاندارد نمودن نمره دو درس مشاهده می‌کنیم که نمره وی در درس فیزیک بالاتر از درس ریاضی می‌باشد به عبارتی علی درس فیزیک را بهتر از درس ریاضی امتحان داده است.

با سپاس فراوان از توجه شما عزیزان

موفق باشید