

به نام خداوند بخشنده مهربان

آمار و روش تحقیق

مجموعه:

(روانشناسی)

مؤلف:

دکتر جعفر میببی

آمارهای آزمون دکتری

دکتر حبیبی، جعفر
 آمار و روش تحقیق / دکتر جعفر حبیبی
 تهران - مشاوران صعود ماهان: ۱۳۹۹
 ۱۸۲ص: جدول، نمودار، (آمادگی آزمون دکتری)
 شابک
 وضعیت فهرست نویسی: فیبا مختصر
 فارسی - چاپ اول
 ۱- آمار و روش تحقیق
 ۲- آزمون دوره های تحصیلات تکمیلی
 ج - عنوان
 شماره کتابشناسی ملی

ISBN/N: 978-600-458-700 -6



انتشارات مشاوران صعود ماهان



- ☐ نام کتاب: آمار و روش تحقیق
- ☐ مدیران مسئول: مجید و هادی سیاری
- ☐ مولف: دکتر جعفر حبیبی
- ☐ مدیر برنامه ریزی و تولید محتوا: سمیه بیگی
- ☐ ناشر: مشاوران صعود ماهان
- ☐ نوبت و تاریخ چاپ: اول / ۱۳۹۹
- ☐ تیراژ: ۱۰۰۰ نسخه
- ☐ قیمت: ۹۴۰/۰۰۰ ریال
- ☐ شابک: ISBN ۹۷۸-۶۰۰-۴۵۸-۷۰۰-۶

انتشارات مشاوران صعود ماهان: تهران - خیابان ولیعصر، بالاتر از تقاطع ولیعصر مطهری، پلاک ۲۵۰

تلفن: ۸۸۱۰۰۱۱۳ و ۸۸۴۰۱۳۱۳

کلیه حقوق مادی و معنوی این اثر متعلق به موسسه آموزش عالی آزاد ماهان می باشد. و هرگونه اقتباس و کپی برداری از این اثر بدون اخذ مجوز پیگرد قانونی دارد.

بنام خدا

ایمان داریم که هر تغییر و تحول بزرگی در مسیر زندگی بدون تحول معرفت و نگرش میسر نخواهد بود. پس بیایید با اندیشه توکل، تفکر، تلاش و تحمل در توسعه دنیای فکریمان برای نیل به آرامش و آسایش توأمان اولین گام را برداریم. چون همگی یقین داریم دانایی، توانایی می‌آورد

شاد باشید و دلی را شاد کنید

برادران سیاری

آمار و روش تحقیق یکی از دروس تأثیرگذار برای کسب موفقیت در مقاطع ارشد و دکتری بسیاری از رشته‌ها از جمله روان‌شناسی و علوم تربیتی است. این درس دارای سرفصل‌هایی مشترک در تمامی گرایش‌های روان‌شناسی و علوم تربیتی بوده و در چند سال اخیر، سطح سؤالات کنکورهای ارشد و دکتری نیز یکسان و سؤالات از مباحثی مشترک طراحی شده است.

در کتاب حاضر سعی بر آن شده است که مطالب آماری و پژوهشی به‌طور جامع، با بیانی روان و با استفاده از تجربه سال‌ها تدریس در دروس ریاضی، آمار و روش‌های تحقیق، خدمت شما عزیزان ارائه شود که هم قابلیت استفاده برای آمادگی کنکورهای ارشد و دکتری را داشته باشد و هم بتواند به‌عنوان کتابی راهگشا در کنار منابع اصلی در دوره‌های ارشد و دکتری رشته‌های روان‌شناسی و علوم تربیتی مورد استفاده قرار گیرد.

آشنایی با روش‌های مختلف آماری و انواع روش‌های تحقیق، تنها به مرحله آمادگی کنکور محدود نشده، بلکه دانشجویان عزیز در دروس دانشگاهی در مقاطع مختلف، پایان‌نامه ارشد، آزمون جامع دکتری و همچنین در رساله دکتری، نیاز به تسلط بر آمار و تکنیک‌های تحقیق خواهند داشت. اما غالباً به دلایل مختلفی مانند ضعف پایه ریاضی، بیان نامفهوم کتاب‌ها، عدم تسلط کافی برخی از اساتید و ... یادگیری کامل، مفهومی و کاربردی این درس به مشکلی بزرگ برای اغلب دانشجویان علوم انسانی تبدیل شده است. لذا تلاش شده است نکات آماری و پژوهشی، به‌صورت مفهومی و با رویکردی کاربردی خدمت شما عزیزان ارائه شود.

ترتیب ارائه مطالب در کتاب حاضر، کمی متفاوت با کتب دیگر (منابع اصلی و کمک‌آموزشی) بوده و سعی بر آن شده تا بهترین نظم ممکن را برای یادگیری فراهم سازد. به‌طور خاص در قسمت روش‌های تحقیق، مباحث به‌گونه‌ای مطرح شده است که دانشجو بعد از آشنایی با مقدمات تحقیق، با دسته‌بندی کلی پژوهش‌های کمی و کیفی در ابتدا آشنایی کامل و عمیق پیدا کرده و سپس به ترتیب از ساده‌ترین تا پیچیده‌ترین روش تحقیق مورد بحث قرار بگیرد. به‌عنوان مثال روش تحقیق آزمایشی که به دقیق‌ترین روش پژوهشی معروف است، در آخرین فصل روش‌های تحقیق مورد بحث قرار گرفته است.

اگر بخواهیم مطابقت با محتوای کنکورهای ارشد و دکتری را مدنظر قرار دهیم، این کتاب دارای پوشش حداقل ۸۰ درصدی بوده و در کنار درسنامه، کتاب تست با پاسخنامه تشریحی کنکورهای ارشد و دکتری سال‌های اخیر نیز تدارک دیده شده تا نیاز عزیزان را در مسیر آمادگی این آزمون‌ها تا حدود زیادی به‌طرف سازد. لذا توصیه می‌شود در کنار این دو مجموعه کتاب و تست، اگر دانشجویان عزیز تسلط کافی پیدا کردند و علاقه به مطالعه گسترده‌تر داشتند، صرفاً از منابع اصلی (و نه کمک‌آموزشی) مانند

کتاب دلاور، فرگوسن، شیولسون، میرز و ... استفاده کنند.

در پایان شکر خالق را به‌جا آورده و از خداوند حکیم مسئلت دارم تا این توفیق را نصیب بنده کرده تا بتوانم برای دانش‌پژوهان عزیز، معلمی سالم، آگاه و راهنما باشم.
جعفر حبیبی - مدرس آمار، روش‌های تحقیق و نرم‌افزارهای آماری

فهرست مطالب

بخش اول - آمار

فصل اول - شاخص‌های توزیع فراوانی و نمودار.....	۶
فصل دوم - شاخص مرکزی	۱۵
فصل سوم - شاخص‌های پراکندگی.....	۲۳
فصل چهارم - نمرات استاندارد و منحنی طبیعی.....	۳۰
فصل پنجم - همبستگی	۳۶
فصل ششم - پیش‌بینی	۴۴
فصل هفتم - برآورد (تعمیم)	۴۹
فصل هشتم - آزمون فرضیه.....	۵۴
فصل نهم - آزمون‌های آماری.....	۵۹
فصل دهم - آزمون‌های پارامتریک.....	۶۷
فصل یازدهم - آزمون مقایسه چندگروهه.....	۷۳
فصل دوازدهم - همبستگی چندمتغیره و انواع رگرسیون.....	۸۹
فصل سیزدهم - احتمالات.....	۹۳
جداول توزیع‌های آماری.....	۹۹

بخش دوم - روش تحقیق

فصل اول - متغیرها.....	۱۱۸
فصل دوم - تحقیقات و منابع شناخت	۱۲۲
فصل سوم - موضوع، سوال و فرضیه‌های تحقیق	۱۳۰
فصل چهارم - ابزارهای اندازه‌گیری و طبیف‌های سنجش.....	۱۳۶
فصل پنجم - نمونه‌گیری.....	۱۴۷
فصل ششم - تحقیقات کیفی.....	۱۵۱
فصل هفتم - روش‌های تحقیق کیفی.....	۱۵۷
فصل هشتم - روش‌های تحقیق کمی از نوع توصیفی.....	۱۶۶
فصل نهم - روش‌های تحقیق کمی از نوع تجربی.....	۱۷۵

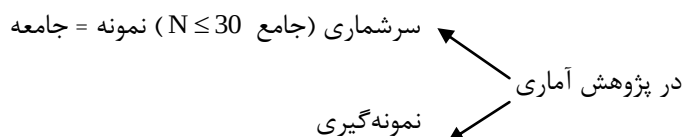
بخش اول

آمار

فصل اول

شاخص‌های توزیع فراوانی و نمودارها

تعریف آمارهای توصیفی و استنباطی



- در پژوهش‌های سرشماری آمار توصیفی کفایت می‌کند.
- آمار توصیفی نمونه را توصیف می‌کند.
- در تحقیقاتی که نمونه‌گیری انجام می‌شود، بعد از توصیف نمونه نتایج آن را به جامعه تعمیم می‌دهیم و از آمار استنباطی استفاده می‌کنیم.

آمار توصیفی: طبقه‌بندی - تلخیص و توصیف اطلاعات حاصل از نمونه

۱. جدول توزیع فراوانی
 ۲. نمودار
 ۳. شاخص‌های مرکزی (مد / میانه / میانگین)
 ۴. شاخص‌های پراکندگی
 ۵. نمرات استاندارد و منحنی طبیعی
 ۶. همبستگی و رگرسیون
- مهم‌ترین کارکرد آمار توصیفی چیست؟ تلخیص (خلاصه کردن)
 - چگونه به هدف می‌رسد؟ داده‌های عددی را در شاخص‌های آماری مثل میانگین و واریانس خلاصه می‌کند.
 - هدف آمار توصیفی: توصیف اطلاعات حاصل از نمونه

آمار استنباطی: به تحلیل و تفسیر و تعمیم نتایج حاصل از تنظیم و محاسبه مقدماتی آماری تکیه دارد.

هدف آمار استنباطی: تعمیم اصول و یافته‌ها برای پیش‌بینی حوادث

آمار استنباطی (تحلیل آماری) به دو دسته آزمون فرضیه و تعمیم تقسیم می‌شود:

۱. آزمون فرضیه

آزمون‌های آماری مانند تحلیل واریانس وظیفه آزمودن فرضیه و بررسی معناداری را دارند؛ در واقع تأیید یا رد فرضیه‌ها با مشخص می‌کنند. به دلایل مختلفی، آزمون‌ها از گستردگی و تعدد بالایی برخوردارند. از مهم‌ترین این دلایل می‌توان به انواع فرضیه (همبستگی، مقایسه‌ای و علی) و انواع سطح سنجش متغیرها (اسمی، ترتیبی، فاصله‌ای و نسبی) اشاره کرد.

۲. تعمیم: برآورد ویژگی‌های جامعه از روی ویژگی‌های نمونه

حروف یونانی حروف لاتین

به ویژگی‌های عددی جامعه، پارامتر و به ویژگی‌های عددی نمونه، آماره گفته می‌شود.

جدول ۱ علائم مربوط به پارامتر و آماره

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

وجه مشترک آمار توصیفی و استنباطی: شاخص نمونه (از شاخص آماری نمونه جهت برآورد پارامتر جامعه استفاده می‌کنند) برآوردکننده: ویژگی نمونه‌ای که به صورت تصادفی از جامعه انتخاب شده باشد.

حدود واقعی اعداد

طبق قرارداد علم آمار، در جهت سنجش و اندازه‌گیری دقیق‌تر، می‌بایست هر داده کمی را به صورت طیفی پیوسته از حد پایین تا حد بالا در نظر بگیریم (نیم واحد پایین‌تر تا نیم‌واحد بالاتر).

۱۲	۱۲/۷۴۰	۱۲/۷۰
۱۲/۵	۱۲/۷۴۵	۱۲/۷۳۵
۱۱/۵	۱۲/۷۴۵	۱۲/۷۳۵
۱۲/۶۵	۱۲/۷۳۵	۱۲/۷۰

جدول توزیع فراوانی

۱. جدول غیرطبقاتی

نمره خام X	فراوانی f
۱۵	۲
۱۲	۳
۵	۱

۲. جدول طبقاتی (دارای خطای طبقه‌بندی)^۱

X	f
۵ - ۰	۲
۱۰ - ۵	۴

^۱ Error of grouping

زمانی که تعداد داده‌ها و همچنین پراکندگی آن‌ها زیاد باشد از جدول طبقاتی استفاده می‌کنیم. (خطای طبقه‌بندی عبارت است از اختلاف بین شاخص آماری محاسبه شده در استفاده از توزیع فراوانی طبقه‌بندی شده و توزیع فراوانی طبقه‌بندی نشده)

نکته: خطای طبقه‌بندی با طول طبقات رابطه مستقیم و با تعداد طبقات رابطه عکس دارد.

خطای بیشتر B	A
۲۵ - ۰	۱۰ - ۰
۵۰ - ۲۵	۲۰ - ۱۰
⋮	⋮
	۱۰۰ - ۹۰

معرفی شاخص‌های توزیع فراوانی^۱

$$R = \text{Max} - \text{Min} + 1$$

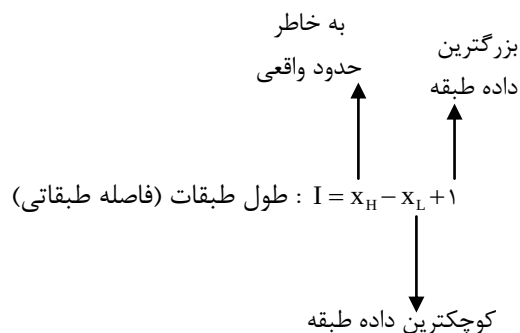
↓

به خاطر حدود واقعی

$$۶, ۹, ۱۰, ۱۵, ۲۶$$

$$R = ۲۶ / ۵ - ۵ / ۵ = ۲۱$$

$$R = ۲۶ - ۶ + ۱ = ۲۱$$



x	f
۸-۱۳	
۱۴-۱۹	

$$i = ۱۳ - ۸ + ۱ = ۶$$

$$k = \frac{R}{I}$$

تعداد طبقات به کمک حاصل تقسیم دامنه تغییرات بر طول طبقات می‌توان تعداد طبقات جدول را محاسبه کرد.

نحوه ساختن توزیع فراوانی طبقه‌بندی شده

اولین مرحله: محاسبه دامنه تغییرات

دومین مرحله: تعیین تعداد طبقات

^۱ Frequency distribution



معمولاً تعداد طبقات بین ۱۰ تا ۲۰ طبقه اختیار می‌شود، اما می‌توان از فرمول زیر که به قانون استرژ معروف است نیز استفاده کرد:

$$K = 1 + 3 / 3 \log(n)$$

در فرمول بالا n و K به ترتیب عبارت‌اند از تعداد اعداد، و طبقه‌ها، پایه لگاریتم بر مبنای ۱۰ است. به طور مثال، چنانچه تعداد کل اعداد جمع‌آوری شده مساوی ۶۰ باشد، تعداد طبقه‌ها فرمول مورد بحث، برابر است با:

$$(\log^{10} \text{ در پایه } ۱۰ \text{ مساوی } ۱/۷۷۸۲ \text{ است})$$

$$K = 1 + 3 / 3 \log n$$

$$K = 1 + 3 / 3 \log^{10}$$

$$K = 1 + 3 / 3 (1/۷۷۸۲)$$

$$K = ۶/۸۶۸۰۶ \approx ۷$$

سومین مرحله: تعیین طول طبقات

$$I = \frac{R}{K}$$

نمونه‌ای از یک جدول توزیع فراوانی غیرطبقاتی:

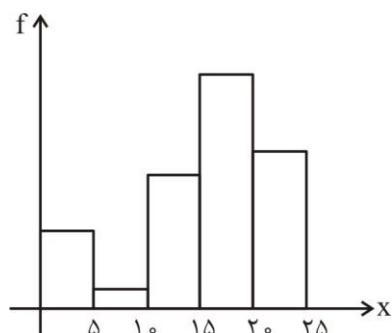
x	f (فراوانی مطلق)	Cf (فراوانی تراکمی / تجمعی)	$\frac{f}{n}$ فراوانی نسبی	$\frac{f}{n} \times ۱۰۰$ فراوانی درصدی	$\frac{Cf}{n}$ فراوانی تراکمی نسبی	$\frac{Cf}{n} \times ۱۰۰$ فراوانی تراکمی درصدی
۵	۲	۲	۰/۲	۲۰٪	۰/۲	۲۰٪
۸	۱	۳	۰/۱	۱۰٪	۰/۳	۳۰٪
۹	۳	۶	۰/۳	۳۰٪	۰/۶	۶۰٪
۱۳	۱	۷	۰/۱	۱۰٪	۰/۷	۷۰٪
۱۶	۳	۱۰	۰/۳	۳۰٪	۱	۱۰۰٪

$$\sum f = N$$

نکته مهم در تنظیم جدول توزیع فراوانی این است که اگر داده‌ها یا طبقات از بزرگ به کوچک مرتب شوند، فراوانی تراکمی یا تجمعی از سمت پایین به بالای جدول محاسبه خواهد شد؛ در واقع فراوانی تراکمی همواره از سمت عدد یا طبقه کوچک شروع می‌شود.

«نمودارها» ← نمایش گرافیکی داده‌ها (ساده‌سازی)

۱. نمودار هیستوگرام^۱



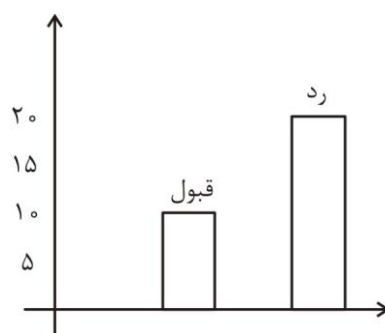
x	f
۰-۵	۴
۵-۱۰	۱
۱۰-۱۵	۷
۱۵-۲۰	۱۱
۲۰-۲۵	۸

- این نمودار در مقیاس‌های فاصله‌ای و نسبی (کمی و پیوسته) استفاده شده و کاربرد مهم آن در آمار، بررسی تقارن توزیع است.

۲. نمودار ستونی^۲ (میله‌ای)

جدول ۲-۱ توزیع فراوانی نمره‌های ۲۵ نفر در یک آزمون فیزیک

درجه‌بندی	فراوانی (f)	درصد (p)
قبول	۱۰	۴۰
رد	۱۵	۶۰



شکل ۲-۲ نمودار ستونی نمره‌های آزمون فیزیک

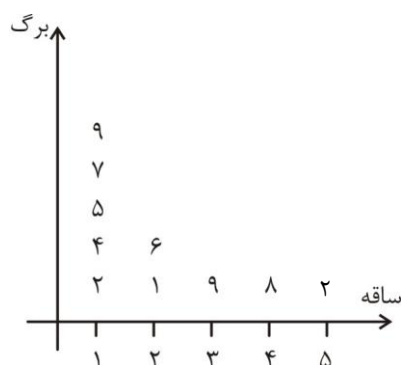
کاربرد آن همانند هیستوگرام است (بررسی تقارن توزیع) با این تفاوت که در سطح داده‌های اسمی کاربرد دارد.

^۱ histogram

^۲ bargraph

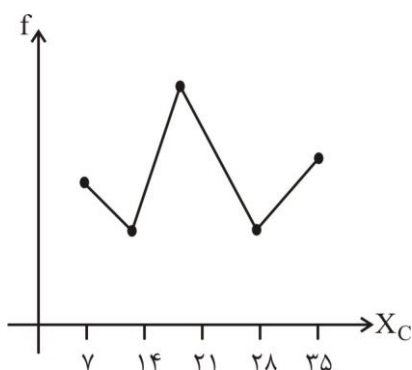
۳. نمودار ساقه و برگ (شاخه و برگ)^۱

۲۶ ۵۲ ۱۲ ۳۹ ۱۷
۱۴ ۴۸ ۱۹ ۲۱ ۱۵



- از نظر مقیاس و کاربرد این نمودارها همانند هیستوگرام هستند، با این تفاوت که اگر تعداد داده‌ها کم باشد اطلاعات جزئی‌تری در اختیار قرار می‌دهد.

۴. نمودار چندضلعی



نقطه میانی طبقه $x_c = \frac{x_L + x_H}{2}$ یا میانگین طبقه ↓

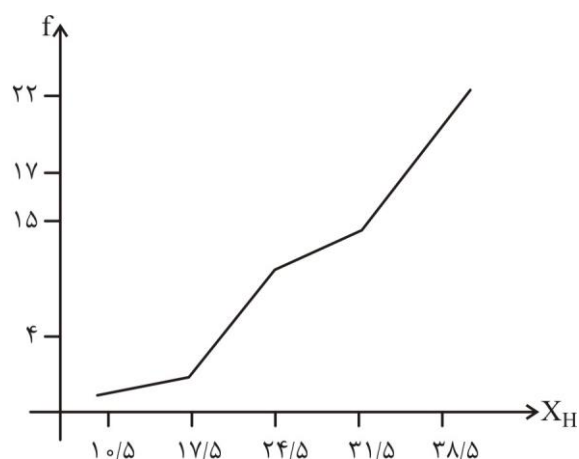
X	f	x_c
۱۰-۴	۴	۷
۱۷-۱۱	۱	۱۴
۲۴-۱۸	۱۰	۲۱
۳۱-۲۵	۲	۲۸
۳۸-۳۲	۵	۳۵

این نمودار نیز برای داده‌های کمی و پیوسته استفاده می‌شود و پرکاربردترین نمودار به شمار می‌رود، زیرا جهت مقایسه دو یا چند توزیع از آن استفاده می‌شود.

^۱ Stem and leaf plot

۵. نمودار چندضلعی تراکمی (اجایو)^۱

x	f	cf	X _H
۱۰-۴	۴	۴	۱۰/۵
۱۷-۱۱	۱	۵	۱۷/۵
۲۴-۱۸	۱۰	۱۵	۲۴/۵
۳۱-۲۵	۲	۱۷	۳۱/۵
۳۸-۳۲	۵	۲۲	۳۸/۵



این نمودار کمتی و پیوسته است و جهت نمایش وضعیت یک داده نسبت به سایر داده‌ها (گروه) استفاده می‌شود.

۶. نمودار دایره‌ای^۲

x	f
بالینی	۲۰
عمومی	۶۰
سلامت	۴۰

$$\frac{f}{n} \times ۳۶۰ = \text{اندازه قطاع}$$

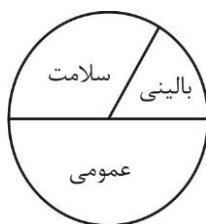
$$\frac{۲۰}{۱۲۰} \times ۳۶۰ = ۶۰ \text{ بالینی}$$

$$\frac{۶۰}{۱۲۰} \times ۳۶۰ = ۱۸۰ \text{ عمومی}$$

$$\frac{۴۰}{۱۲۰} \times ۳۶۰ = ۱۲۰ \text{ سلامت}$$

^۱ Cumulative (Ogive)

^۲ Pie chart

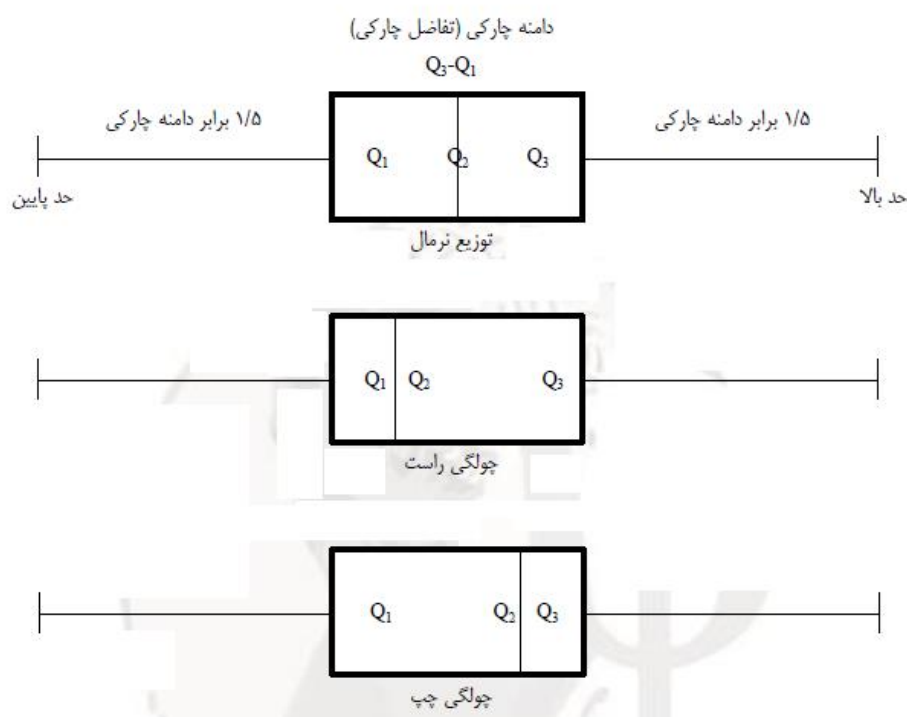


نمودار دایره‌ای مقیاس اسمی - ترتیبی است و جهت نمایش نسبت جزء به کل استفاده می‌شود.

۷. نمودار جعبه‌ای^۱

مهمترین کاربرد نمودار جعبه‌ای، یافتن داده پرت است.

به طوری که $1/5$ برابر دامنه چارکی محاسبه و از مقدار چارک اول کاسته و به مقدار چارک سوم افزوده می‌شود. دامنه به دست آمده دامنه نرمال بوده و اعداد بیرون از این دامنه داده پرت به شمار می‌روند.



¹ Box plot

فصل دوم

شاخص‌های مرکزی

۱-مد^۱ یا نما M_O

مد شاخصی مبتنی بر تکرار است و به طبقه‌ای اطلاق می‌شود که دارای بیشترین فراوانی باشد. این شاخص ضعیف‌ترین و بی-ثبات‌ترین شاخص مرکزی به شمار می‌رود، زیرا حتی ویژگی ترتیب برای آن مهم نیست (مقیاس اسمی) و لزوماً در مرکز توزیع قرار نمی‌گیرد. از مد می‌توان در تمامی مقیاس‌ها استفاده کرد و کاربرد آن زمانی است که قصد محاسبه سریع و آسان یک شاخص مرکزی را داشته باشیم.

محاسبات نما:

زمانی که دو عدد یا دو طبقه متوالی، دارای بیشترین فراوانی و به مقدار یکسان باشند، میانگین این دو عدد یا طبقه، به عنوان مد یا نما معرفی خواهد شد.

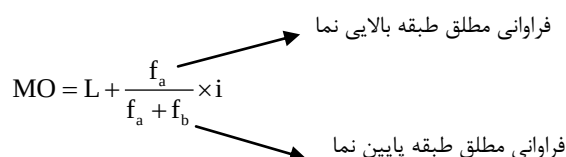
طرز محاسبه مد در داده‌های طبقه‌بندی شده (پیوسته)

برای محاسبه مد مراحل زیر را به ترتیب انجام می‌دهیم:

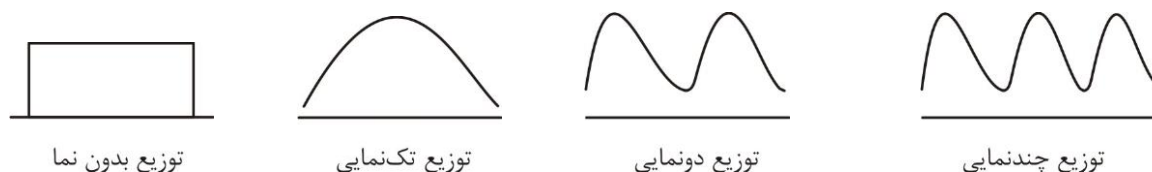
۱- ابتدا طبقه‌ای که بیشترین فراوانی مطلق را دارد انتخاب می‌کنیم.

۲- از فرمول مقابل مد را محاسبه می‌کنیم:

$$M_O = L + \frac{f_a}{f_a + f_b} \times i$$



نکته: توزیع‌های آماری مورد توجه در این کتاب، توزیع‌های تک‌نمایی هستند؛ اما لازم به ذکر است که توزیع فراوانی داده‌ها دونمایی یا چندنمایی باشد:



۲-میان^۲ M_d

میان^۲ نقطه وسط یا پنجاه درصدی داده‌ها است.

این شاخص مبتنی بر ترتیب و دارای مقیاس ترتیبی است؛ پس در مقیاس‌هایی می‌توان از آن استفاده کرد که ویژگی ترتیب را داشته باشند؛ یعنی مقیاس‌های ترتیبی، فاصله‌ای و نسبی.

¹ mode
² median

میانه برخلاف میانگین شاخصی کیفی است که تحت تأثیر ارزش عددی یا حجم داده‌ها قرار نمی‌گیرد. لذا میانه به داده‌ها پرت کوچک یا بزرگ حساس نیست؛ بنابراین می‌توان گفت اگر توزیع داده‌ها چولگی یا داده پرت داشته باشد، بهترین و مناسب‌ترین شاخص مرکزی میانه است.

$$\sum |x - m_d| \leq \sum |x - A|$$

همواره و در هر توزیعی مجموع قدر مطلق انحراف داده‌ها از میانه حداقل است.

«محاسبه میانه»

۱. وقتی تعداد داده‌ها فرد باشد.

۲. وقتی تعداد داده‌ها زوج باشد.

۱۸، ۱۵، ۱۰، ۹، ۵

۵، ۸، ۱۰، ۱۴، ۱۷، ۲۰

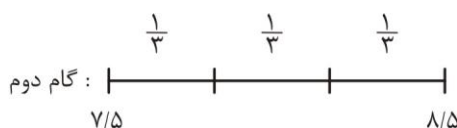


از این دو میانگین می‌گیریم (۱۲)

۳. زمانی که در میانه توزیع داده تکراری داشته باشیم.

Md
۵، ۷، ۸، ۸، ۱۲، ۱۵، ۲۰

مکان میانه: $\frac{n}{2} = \frac{8}{2} = 4$ گام اول



$$Md = 7/5 + \frac{2}{3} = 8/5$$

$$Md = \text{حد پایین عدد تکراری} + \frac{\text{تعداد تکرارهای قبل از میانه}}{\text{تعداد کل تکرارها}}$$

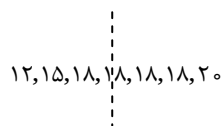
Md
۱۵ ۱۹ ۱۹ ۱۹ ۲۶ ۲۶ ۲۶ ۲۶ ۳۲ ۳۲

مکان میانه: $\frac{10}{2} = 5$

$$Md = 25/5 + \frac{1}{4} = 25/4$$

x	f
۱۵	۱
۱۹	۳
۲۶	۴
۳۲	۲

Md



$$\text{مکان میانه} : \frac{n}{2} = \frac{7}{2} = 3.5$$

$$Md = 17.5 + \frac{1}{4}$$

۴. زمانی که تعداد داده‌ها زیاد باشد یا جدول طبقه‌بندی شده داشته باشیم:

$$Md = L + \frac{\frac{h}{2} - cf}{f} \quad (i)$$

L = پایین حد واقعی طبقه‌ای که میانه در آن قرار دارد.

cf = فراوانی تراکمی طبقه قبل از طبقه میانه

f = فراوانی مطلق طبقه‌ای که میانه در آن قرار دارد.

i = فاصله طبقات

h = تعداد نمره‌ها

به طور مثال:

جدول توزیع فراوانی نمره‌های یک آزمون

طبقات	f	cf
۴۸-۵۰	۲	۵۰
۴۵-۴۷	۳	۴۸
۴۲-۴۴	۴	۴۵
۳۹-۴۱	۶	۴۱
۳۶-۳۸	۸	۳۵
۳۳-۳۵	۸	۲۷
۳۰-۳۲	۷	۱۹
۲۹-۲۷	۶	۱۲
۲۴-۲۶	۳	۶
۲۱-۲۳	۲	۳
۱۸-۲۰	۱	۱

محاسبه میانه اعداد طبقه‌بندی شده در شرایطی که فاصله طبقه‌ها بزرگ‌تر از یک است.

محاسبه میانه برای توزیع فراوانی زیر شامل پنج مرحله به شرح زیر است:

۱. تقسیم تعداد کل نمره‌ها بر ۲

$$50 \div 2 = 25$$

۲. مراجعه به ستون فراوانی تراکمی و پیدا کردن اولین طبقه‌ای که فراوانی تراکمی آن مساوی یا بزرگ‌تر از $\frac{N}{2}$ یعنی ۲۵ است. با مراجعه به جدول اولین طبقه‌ای است که فراوانی تراکمی آن بزرگ‌تر از ۲۵ است (برای پیدا کردن این طبقه از پایین‌ترین طبقه شروع می‌کنیم).

۳. حد پایین واقعی این طبقه یا L را تعیین می‌کنیم. $L = 32/5$

۴. مقادیر cf و f را از جدول توزیع فراوانی استخراج می‌کنیم (به ترتیب ۱۹ و ۸).

۵. مقادیر تعیین‌شده به شرح فوق را در فرمول زیر جایگزین می‌کنیم.

$$Md = 32/5 + \frac{\frac{50}{2} - 19}{8} (3) = 34/75$$

۱. میانگین^۱ [حسابی] M یا $\bar{x}$

ویژگی‌های میانگین

- ۱- میانگین دارای مقیاس فاصله‌ای است و در مقیاس‌های فاصله‌ای و نسبی می‌توان از آن استفاده کرد.
- ۲- اگر توزیع داده‌ها نرمال باشد بهترین و مناسب‌ترین شاخص مرکزی میانگین است.
- ۳- همواره و در هر توزیعی مجموع انحراف داده‌ها از میانگین برابر صفر است، به همین دلیل است که میانگین را مرکز ثقل یا نقطه تعادل توزیع می‌نامند.

$$X = 2 \quad 3 \quad 5 \quad 14 \quad \bar{x} = 6$$

$$X - \bar{x} = \begin{array}{cccc} -4 & -3 & -1 & 8 \\ & \text{---} & & \\ & -8 & & +8 \end{array}$$

$$\sum (x - \bar{x}) = 0 \text{ همواره}$$

۴- همواره و در هر توزیعی مجموع مجذور انحراف داده‌ها از میانگین حداقل است:

$$\sum (x - \bar{x})^2 \leq \sum (x - a)^2 \rightarrow a$$

هر عدد دلخواهی:

محاسبه میانگین:

$$\bar{x} = \frac{\sum x}{n}$$

$$\bar{x} = \frac{\sum fx}{n} \text{ جدول غیرطبقه‌ای}$$

x	f
۱۵	۳
۱۶	۴
۱۰	۲

¹ mean

جدول طبقاتی : $\bar{x} = \frac{\sum fx_c}{n}$

x	f	X_c
۱۵-۲۳	۳	۱۹
۲۴-۳۲	۱	۲۸
۳۳-۴۱	۴	۳۷

مثال: کلاس ۱۰ نفره‌ای دارای میانگین ۱۵ و کلاس ۴۰ نفره‌ای دارای میانگین ۱۰

$$\bar{X} = 15 \quad \bar{X} = 10$$

$$n = 10 \quad n = 40$$

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + \dots}{n_1 + n_2 + \dots} \quad \text{میانگین وزنی - (میانگین کل، میانگین مرکب، میانگین میانگین‌ها)}$$

$$\text{میانگین کل} = \frac{10 \times 40 + 15 \times 10}{50} = \frac{550}{50} = 11$$

فرمول تجربی پیرسون (رابطه میان شاخص‌های مرکزی):

$$m_o = 3m_d - 2\bar{x}$$

مثال: در توزیعی با میانگین ۲۰ و نمای ۵، میانه چند است؟

$$m_d = 15$$

میانگین هندسی^۱

برای محاسبه میانگین درصدها، نسبت‌ها یا نرخ‌ها استفاده می‌شود.

$$G = \sqrt[n]{x_1 x_2 \dots x_n}$$

مثال: میانگین هندسی اعداد ۸، ۲۷، ۱۲۵

$$G = \sqrt[3]{8 \times 27 \times 125} = 30$$

میانگین هارمونیک^۲ یا همساز

برای محاسبه میانگین داده‌هایی با واحد ترکیبی (مانند $\frac{m}{s}$) استفاده می‌شود.

$$H \text{ یا } HM = \left(\frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}} \right)$$

نکته: در صورت محاسبه میانگین‌ها برای یک توزیع فراوانی، میانگین حسابی از نظر عددی بزرگ‌تر از میانگین هندسی و هارمونیک است:

$$\bar{x} > G > HM$$

^۱ Geometric mean

^۲ Harmonic mean

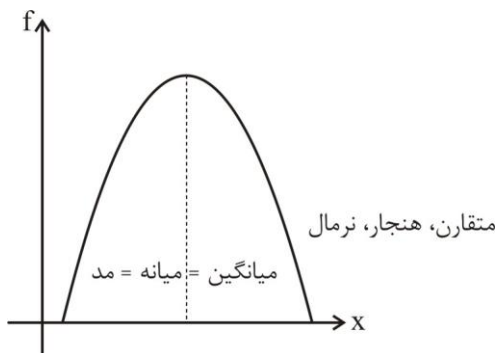
فرمول محاسبه میانگین همساز از طریق میانگین‌های هندسی و حسابی.

$$HM = \frac{G^2}{\bar{X}}$$

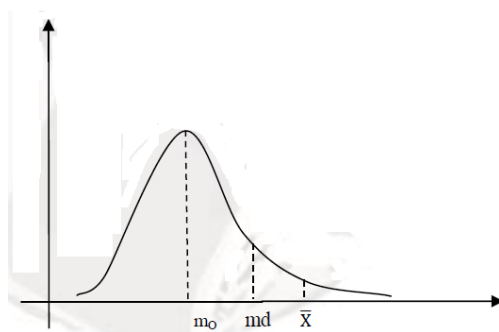
چولگی^۱ (کجی)

شاخص چولگی تحت تأثیر داده پرت است و جهت بررسی تقارن استفاده می‌شود.

x	f
۰	۲
۳	۴
۵	۷
۸	۹
۱۱	۱۵
۱۴	۸
۱۷	۷
۲۰	۱



x	f
۰	۴
۲	۱۷
۳	۱۲
۵	۸
۷	۵
۸	۵
۱۹	۱
۲۰	۱



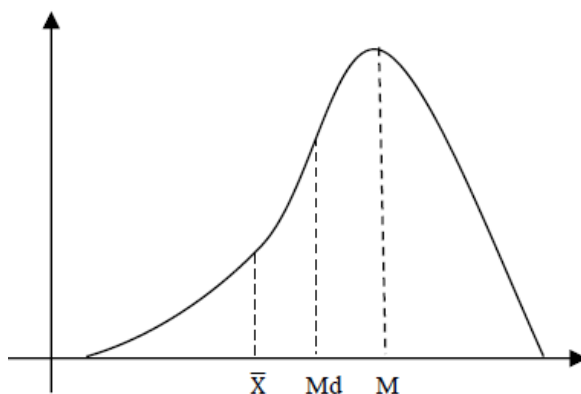
چولگی راست یا مثبت:

با اینکه اکثریت نمره کم دارند، به خاطر چندتا نمره بزرگ میانگین بزرگ‌تر از نما و میانه شده است.

داده پرت بزرگ ← میانگین را بزرگ می‌کند ← چولگی راست

مثال: توزیع درآمد در تهران

x	f
۰	۲
۲	۱
۱۳	۴
۱۴	۶
۱۵	۶
۱۶	۱۰
۱۷	۱۵
۱۸	۶
۲۰	۴



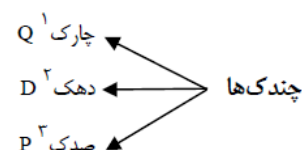
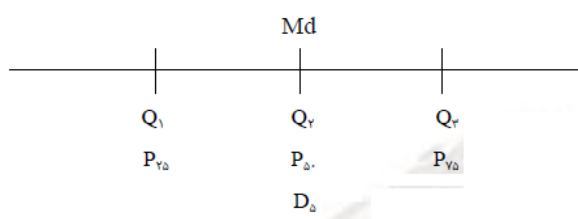
^۱ Skewness

چولگی چپ یا منفی

داده پرت کوچک ← میانگین را کوچک می‌کند ← چولگی چپ

محاسبه چولگی

$$sk = \frac{\bar{x} - m_o}{s} \quad |sk| \leq 1 \quad \text{نرمال}$$



چندک‌ها شاخص‌های وضعی و هم‌خانواده میانه هستند، به طوری که همانند میانه دارای مقیاس ترتیبی بوده و محاسبات این شاخص‌ها نیز بسیار شبیه یافتن میانه است.

از جمله معروف‌ترین شاخص‌های وضعی می‌توان به چارک، دهک و صدک اشاره کرد که به ترتیب، توزیع را به چهار قسمت، ده قسمت و صد قسمت مساوی تقسیم می‌کنند.

مثال: چارک‌ها را در توزیع زیر پیدا کنید.

فرمول محاسبه چندک‌ها:

$$\begin{array}{c} 7, 9, 9 \mid 9, 12, 18 \mid 20, 25, 25 \mid 25, 25, 25 \\ Q_1 = 8.5 + \frac{2}{3} \quad Q_2 = 19 \quad Q_3 = 24.5 + \frac{2}{5} \end{array}$$

همان فرمول محاسبه میانه است، با این تفاوت که به جای n تقسیم بر ۲، باید نسبت مورد نظر را جای‌گذاری کرد. برای مثال محاسبه چارک اول خواهیم داشت:

$$Q_1 = L + \left[\frac{\frac{n}{4} - CF}{F} \right] i$$

یا اگر قصد محاسبه چارک سوم را داشته باشیم، کافی است به جای $\frac{n}{4}$ مقدار $\frac{3n}{4}$ استفاده کنیم.

یا برای محاسبه دهک هفتم: $\frac{7n}{10}$

برای محاسبه نقطه ۸۴ درصدی: $\frac{84n}{100}$

¹ quartile
² decile
³ percentile

فصل سوم

شاخص‌های پراکندگی



شاخص‌های پراکندگی^۱ (فاصله‌ای)



۱. دامنه تغییرات^۲ $R = \max - \min + ۱$

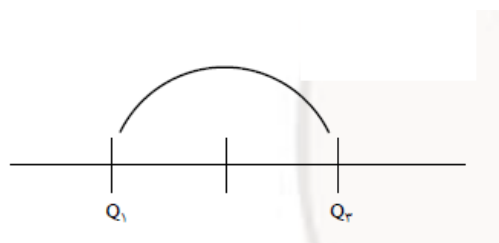
پراکنده‌تر $R = ۱۸$: کلاس A $۱۹...۳,۲$

همگن‌تر $R = ۵$: کلاس B $۱۸...۱۴,۱۴$

نه تنها دامنه تغییرات بلکه تمامی شاخص‌های پراکندگی به دنبال فاصله میان داده‌ها و دارای مقیاس صاله‌ای هستند. دامنه تغییرات ضعیف‌ترین و بی‌ثبات‌ترین شاخص پراکندگی به شمار می‌رود، در محاسبه آن فقط از دو عدد ابتدایی و انتهایی استفاده شده و با تغییر یکی از این اعداد مقدار R نیز شدیداً تغییر می‌کند؛ لذا در میان شاخص‌های پراکندگی، دامنه تغییرات بیشترین حساسیت را به داده پرت دارد. پس زمانی در آمار از این مقیاس استفاده می‌شود که بدون داشتن داده پرت قصد محاسبه سریع و آسان یک شاخص پراکندگی را داشته باشیم.

۲. انحراف چارکی^۳

$$Q = \frac{Q_3 - Q_1}{۲} \quad (\text{چارک متوسط})$$



انحراف چارکی نشان می‌دهد داده‌ها به طور متوسط چقدر از میانه یا مرکز توزیع انحراف دارند. در واقع انحراف چارکی میزان پراکندگی را در ۵۰٪ وسط توزیع بررسی کرده لذا به داده‌های پرت کوچک یا بزرگ حساس نیست. پس می‌توان گفت زمانی که توزیع داده‌ها چولگی قابل ملاحظه‌ای داشته باشد، بهترین و مناسب‌ترین شاخص پراکندگی انحراف چارکی خواهد بود.

^۱ Variability

^۲ Range

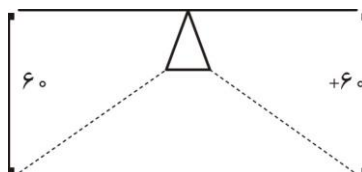
^۳ Quartile deviation

$$3. \text{ انحراف متوسط }^1 AD = \frac{\sum |x - \bar{x}|}{n} \text{ یا MD}$$

انحراف متوسط عبارت است از متوسط قدرمطلق انحراف داده‌ها از میانگین. این شاخص نشان می‌دهد که داده‌ها به طور متوسط چقدر از میانگین انحراف دارند. انحراف متوسط یک ایراد اساسی دارد و آن اینکه به کمک قراردادی به نام قدرمطلق علائم جبری را از بین برده و در عملیات ریاضی و آماری نمی‌توان از آن استفاده کرد.

مثال: در یک توزیع ۳۰ تایی چنانچه مجموع انحرافات منفی از میانگین برابر ۶۰ شده باشد، مقدار انحراف متوسط چند است؟

$$AD = \frac{60 + 60}{30} = \frac{120}{30} = 4$$



$$4. \text{ واریانس }^2 S^2 = \frac{\sum (x - \bar{x})^2}{n}$$

واریانس عبارت است از متوسط مجذور انحراف داده‌ها از میانگین.

$$\text{دلیل نام‌گذاری واریانس: } SS \rightarrow S^2 \quad \text{مجموع مجذورات } \sum (x - \bar{x})^2$$

واریانس پرکاربردترین شاخص پراکندگی است، اما یک ایراد اساسی دارد و آن اینکه به دلیل مجذور کردن انحراف‌ها، اختلاف واحد اندازه‌گیری ایجاد می‌کند.

$$\text{مثلاً } m \rightarrow m^2$$

برای رفع این ایراد است که از واریانس، جذر یا رادیکال گرفته شده و شاخصی به نام انحراف استاندارد ساخته شده است. تفاوت واریانس‌های نمونه و جامعه: در آمار استنباطی (قضیه حد مرکزی) اثبات شده است که واریانس نمونه تصادفی از واریانس کل جامعه کوچک‌تر است (داشتن اریب یا سوگیری). به همین دلیل برای اینکه در محاسبه واریانس نمونه این ایراد را از بین ببریم لازم است در مخرج فرمول از $n-1$ استفاده کنیم.

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

¹ Mean deviation

² Variance

تست: واریانس برآورد (تعمیم) داده‌ها مقابل چقدر است؟

↑ نمونه است و می‌خواهیم تعمیم دهیم.

واریانس نمونه

$$X = 2 \quad 3 \quad 6 \quad 8 \quad 16$$

$$\bar{X} = \frac{35}{5} = 7$$

$(x - \bar{x})$	$(x - \bar{x})^2$	$\sum (x - \bar{x})^2 = 124$
$2 - 7 = -5$	۲۵	
$3 - 7 = -4$	۱۶	
$6 - 7 = -1$	۱	
$8 - 7 = 1$	۱	
$16 - 7 = 9$	۸۱	

$$S^2 = \frac{\sum (x - \bar{x})^2}{n - 1} = \frac{124}{5} = 24.8 \quad \text{نمونه (برآورد) (تعمیم) } 31$$

$$S^2 = \frac{\sum (x - \bar{x})^2}{n} = \frac{124}{5} = 24.8 \quad \text{جامعه حقیقی (واقعی) } 24.8$$

واریانس

$$S^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n - 1} = \frac{369 - \frac{1225}{5}}{4} = \frac{124}{4} = 31 \quad \text{محاسبه واریانس به کمک اعداد خام}$$

$$S = \sqrt{S^2} \quad \text{۵. انحراف استاندارد}^1 \text{ (معیار)}$$

انحراف معیار برابر است با جذر یا رادیکال واریانس یعنی جذر یا ریشه دوم میانگین مجذور انحراف داده‌ها از میانگین. زمانی که توزیع داده‌ها نرمال باشد، یعنی داده پرت یا چولگی شدیدی نداشته باشیم، بهترین و مناسب‌ترین شاخص پراکندگی انحراف استاندارد خواهد بود.

تست: در توزیعی با کجی $SK = -0.73$ چنانچه بهترین شاخص گرایش به مرکز، میانگین انتخاب شده باشد. بهترین شاخص پراکندگی کدام خواهد بود؟

مقدار SK نیز کمتر از یک بوده و در نتیجه توزیع نرمال است.

	مرکزی	پراکندگی
نرمال	$\bar{X}$	S
چولگی	Md	Q

نکته: بین انحراف معیار (S)، انحراف متوسط (MD) و انحراف چارکی همواره رابطه روبه‌رو برقرار است: $Q < MD < S$

¹ Standard deviation

در یک توزیع طبیعی از نمره‌ها، تقریباً ۵۰ درصد داده‌ها در فاصله‌ی $\pm 1Q$ ، ۵۸ درصد در فاصله‌ی $\pm 1MD$ و ۶۸ درصد داده‌ها در فاصله‌های $\pm 1S$ از میانگین قرار دارند. انحراف معیار (S) همواره کمی بیشتر از انحراف متوسط (MD) است.

$$S = \sqrt{S^2 - \frac{i^2}{12}} \quad \text{تصحیح شپرد}^1$$

در جدول طبقاتی، زمانی که میانگین را محاسبه می‌کنیم به دلیل استفاده از x_c مقدار خطا بسیار ناچیز شده و از نظر آماری اهمیتی ندارد اما در محاسبه انحراف استاندارد به دلیل مجوز شدن انحراف‌ها، این خطا افزایش می‌یابد و لازم است به کمک تصحیح شپرد مقدار آن تعدیل شود. ضمناً این تصحیح را زمانی استفاده می‌کنیم که تعداد طبقات کمتر از ۱۲ باشد: $k < 12$

سؤال: در جدول طبقاتی با تعداد ۹ طبقه و طول طبقات ۶ چنانچه واریانس ۲۸ به دست آمده باشد مقدار انحراف استاندارد چقدر است؟

$$S = \sqrt{28 - \frac{36}{12}} = \sqrt{25} = 5$$

$$i = 6$$

$$S^2 = 28$$

$$9 < 12$$

اگر $K > 12$ بود دیگر نیازی به تصحیح نداشت: $S = \sqrt{28}$

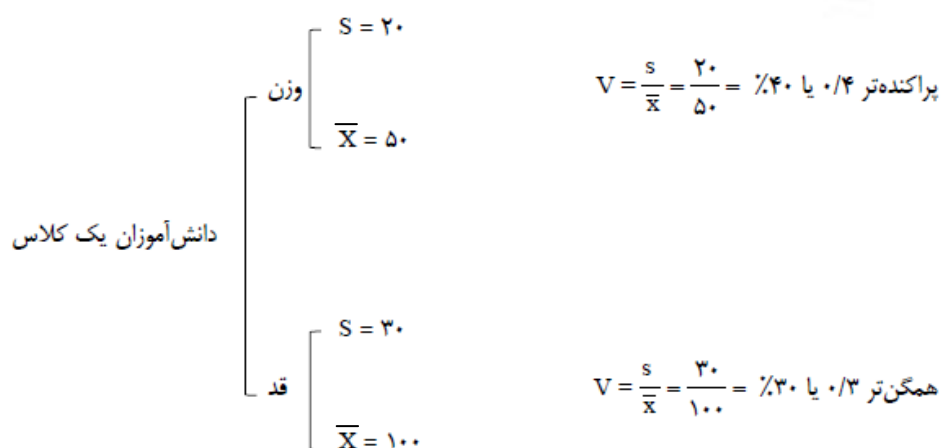
$$\text{ضریب تغییرات}^2 \text{ (ضریب پراکندگی)} \quad v = \frac{S}{\bar{X}} \quad \text{یا} \quad c.v$$

برای مقایسه پراکندگی یک ویژگی در گروه‌های مختلف، می‌توان از S^2 یا S استفاده کرد؛ اما برای مقایسه پراکندگی دو یا چند ویژگی مختلف با واحدهای متفاوت می‌بایست از شاخصی به نام ضریب تغییرات که مستقل از واحد اندازه‌گیری است استفاده کرد.

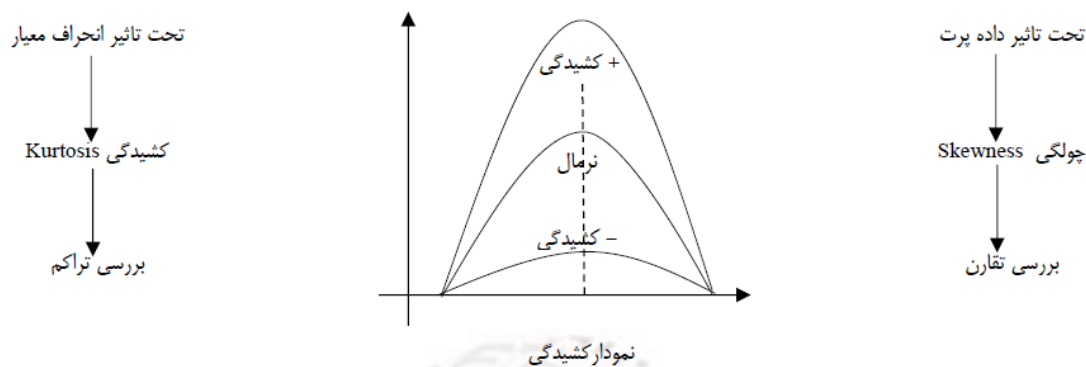
برای مثال:

¹ Sheppard's correction

² Coefficient of variation



کشیدگی^۱



کشیدگی، تحت تأثیر مقدار انحراف استاندارد بوده و جهت بررسی تراکم توزیع از آن استفاده می‌شود؛ به طوری که مقدار کشیدگی، با میزان تراکم توزیع رابطه مستقیم اما با میزان پراکندگی (S) رابطه عکس دارد.
نکته: توزیع نرمال به توزیعی گفته می‌شود که نه چولگی شدید و نه کشیدگی شدید داشته باشد.

گشتاور^۲:

گشتاور مرتبه اول

$$\sum (\bar{x} - x) = 0 : m_1$$

گشتاور مرتبه دوم:

$$\sum (\bar{x} - x)^2 > 0 : m_2$$

$$Ku = \frac{m_4}{S^4} - 3 \quad \text{کشیدگی}$$

$$Sk = \frac{m_3}{S^3} \quad \text{چولگی}$$

¹ Kurtosis
² moment

همان طور که ملاحظه می شود، چولگی (کجی)، تحت تأثیر گشتاور مرتبه سوم بوده و کشیدگی، از گشتاور مرتبه چهارم تأثیر می پذیرد.

تأثیر اعمال ریاضی بر شاخص های مرکزی و پراکندگی:

به طور کلی شاخص های مرکزی به تمامی اعمال ریاضی حساس بوده در حالی که شاخص های پراکندگی فقط به ضرب و تقسیم حساس هستند.

چنانچه تمامی داده های یک توزیع را در عدد ثابتی ضرب یا تقسیم کنیم، شاخص های پراکندگی نیز در همان عدد ضرب یا تقسیم خواهند شد به جز واریانس که در مجذور آن عدد ضرب یا تقسیم می شود.

تست: اگر به بزرگ ترین داده یک توزیع ۱۰ نمره اضافه شود، کدام یک از شاخص های زیر بدون تغییر باقی می ماند؟

- (۱) میانگین (۲) انحراف استاندارد (۳) میانه* (۴) انحراف متوسط
- سؤال: تمامی داده های یک توزیع دو برابر شده، سپس به اضافه ۵ شده و میانگین و واریانس حاصل ۲۵ و ۳۶ به دست آمده است. میانگین و انحراف استاندارد اولیه چه قدر بوده است؟

$$2X + 5$$

$$2\bar{X} + 5 = 25 \quad 2\bar{X} = 20 \quad \bar{X} = 10$$

$$2^2 s^2 = 36 \rightarrow s^2 = 9 : S = 3$$

سؤال: اگر مقدار انحراف معیار $X_1, X_2, \dots, X_n$ برابر ۴ باشد، مقدار واریانس $5X_1 + 10, \dots, 5X_n + 10$ چقدر است؟

$$\begin{array}{c} S = 4 \\ \downarrow \times 5 \\ S_x = 4 \times 5 = 20 \end{array} \quad S = 20 : S^2 = 400$$

مثال: فراوانی تمامی داده ها را در عدد ۳ ضرب کرده ایم، کدام تغییر درباره میانگین اتفاق می افتد؟

- (۱) افزایش (۲) کاهش (۳) تغییر نمی کند* (۴) تغییر می کند

نکته مهم: در صورتی که فراوانی تمام داده ها به شکل یکسانی تغییر کند هیچ شاخص آماری تغییر نمی کند.

پاسخ: کاهش می یابد.

از آنجا که تنها شاخص های مرکزی به جمع و تفریق حساس هستند، پس در صورت اضافه شدن یک عدد به تمامی داده ها، در فرمول ضریب تغییرات مقدار مخرج (میانگین) افزایش پیدا می کند، در حالی که انحراف معیار بدون تغییر باقی می ماند، به همین دلیل مقدار ضریب تغییرات کاهش پیدا خواهد کرد.

فصل چهارم

نمرات استاندارد و منحنی طبیعی