به نام خداوند بخشنده مهربان

آمار و روش تحقیق

مجموعه:

روانشناسی

مؤلف:

دکتر جعفر حبیبی

ارشد و دکتری

آمادگی آزمون ارشد و دکتری

دکتر حبیبی، جعفر
آمار و روش تحقیق / دکتر جعفر حبیبی
تهران ـ مشاوران صعود ماهان: ۱۴۰۴
۱۸۱ص: جدول، نمودار، (آمادگی آزمون ارشد و دکتری)
شابک ISB/N: 978-600-458-700 -6
وضعیت فهرستنویسی: فیپا مختصر
فارسی ـ چاپ چهارم
۱-آمار و روش تحقیق
۲- آزمونها و تمرینها
۲- آزمون دوره‌های تحصیلات تکمیلی
۴-دانشگاه‌ها و مدارس عالی ـ ایران ـ آزمونها
ج ـ عنوان

انتشارات مشاوران صعود ماهان

❐ **نام کتاب:** آمار و روش تحقیق
❐ **مدیران مسئول:** مجید و هادی سیّاری
❐ **مولف:** دکتر جعفر حبیبی
❐ **مدیربرنامه ریزی و تولید محتوا:** سمیه بیگی
❐ **ناشر:** مشاوران صعود ماهان
❐ **نوبت و تاریخ چاپ:** چهارم /۱۴۰۴
❐ **تیراژ:** ۱۰۰۰ نسخه
❐ **قیمت:** ۵/۵۰۰/۰۰۰ریال
❐ **شابک:** ISBN۹۷۸-۶۰۰-۴۵۸-۷۰۰-۶

انتشارات مشاوران صعود ماهان:تهران ـ خیابان ولیعصر، بالاتر از تقاطع ولیعصر مطهری، پلاک ۲۰۵۰
تلفن: ۸۸۱۰۰۱۱۳ و ۸۸۴۰۱۳۱۳

مقدمه ناشر

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

فهرست مطالب

مقدمـــه	
پیشـــگفتار	
بخش اول آمار	۷
فصل اول شاخص‌های توزیع فراوانی و نمودارها	۷
فصل دوم شاخص‌های مرکزی	۱۶
فصل سوم شاخص‌های پراکندگی	۲۴
فصل چهارم نمرات استاندارد و منحنی طبیعی	۳۱
فصل پنجم همبستگی	۳۷
فصل ششم پیش‌بینی	۴۳
فصل هفتم برآورد (تعمیم)	۴۸
فصل هشتم آزمون فرضیه	۵۳
فصل نهم آزمون‌های آماری ناپارامتریک (غیر مشروط)	۵۸
فصل دهم آزمون‌های پارامتریک (مشروط)	۶۶
فصل یازدهم آزمون مقایسه چندگروهه (مقایسه چند میانگین)	۷۲
فصل دوازدهم همبستگی چندمتغیره و انواع رگرسیون	۸۶
فصل سیزدهم احتمالات	۹۲
فصل چهاردهم پیوست	۹۹
بخش دوم روش تحقیق	۱۱۸
فصل اول متغیرها	۱۱۹
فصل دوم تحقیقات و منابع شناخت	۱۲۳
فصل سوم موضوع، سؤال و فرضیه‌های تحقیق	۱۳۱
فصل چهارم ابزارهای اندازه‌گیری و طیف‌های سنجش	۱۳۶
فصل پنجم نمونه‌گیری	۱۴۵
فصل ششم تحقیقات کیفی	۱۴۹
فصل هفتم روش‌های تحقیق کیفی	۱۵۶
فصل هشتم روش‌های تحقیق کمّی [از نوع توصیفی]	۱۶۴
فصل نهم روش‌های تحقیق کمّی [از نوع تجربی]	۱۷۳

پیشگفتـــــــــــــار

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 تجربهٔ 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 مرحلهٔ 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 کمّی 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

Instagram.com/Habibi_Res_Sta , t.me/Habibi_Res_Sta

بخش اول

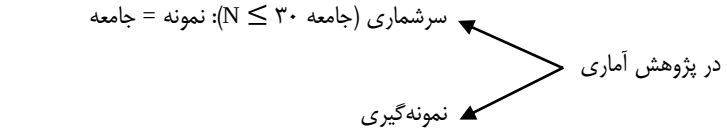
آمار

فصل اول شاخص‌های توزیع فراوانی و نمودارها
فصل دوم شاخص‌های مرکزی
فصل سوم شاخص‌های پراکندگی
فصل چهارم نمرات استاندارد و منحنی طبیعی
فصل پنجم همبستگی
فصل ششم پیش‌بینی
فصل هفتم برآورد (تعمیم)
فصل هشتم آزمون فرضیه
فصل نهم آزمون‌های آماری ناپارامتریک (غیر مشروط)
فصل دهم آزمون‌های پارامتریک (مشروط)
فصل یازدهم آزمون مقایسه چندگروهه (مقایسه چند میانگین)
فصل دوازدهم همبستگی چندمتغیره و انواع رگرسیون
فصل سیزدهم احتمالات
فصل چهاردهم پیوست

فصـــــــــل اول

شاخص‌های توزیع فراوانی و نمودارها

تعریف آمارهای توصیفی و استنباطی



ـ در پژوهش‌های سرشماری آمار توصیفی کفایت می‌کند.

ـ آمار توصیفی نمونه را توصیف می‌کند.

ـ 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؛ درواقع تأیید یا رد فرضیه‌ها را 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.

علائم مربوط به پارامتر و آماره

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

وجه مشترک آمار توصیفی و استنباطی: شاخص نمونه (از شاخص آماری نمونه جهت برآورد پارامتر جامعه استفاده می‌کنند)

برآوردکننده: ویژگی نمونه‌ای که به‌صورت تصادفی از جامعه انتخاب شده باشد.

در واقع به کمک آماره ها ابتدا نمونه‌ی انتخاب شده از جامعه توصیف شده و سپس برآورد و فرض آزمایی پیرامون ویژگی های جامعه صورت می گیرد.

حدود واقعی اعداد

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

۱۲	۱۲/۷۴۰	۱۲/۷۰
۱۱/۵ ۱۲/۵	۱۲/۷۳۵ ۱۲/۷۴۵	۱۲/۶۵ ۱۲/۷۵

جدول توزیع فراوانی

۱. جدول غیرطبقاتی

فراوانی f	نمره خام x
۲	۱۵
۳	۱۲
۱	۵

۲. جدول طبقاتی (دارای خطای طبقه‌بندی[1])

f	x
۲	۵–۰
۴	۱۰–۵
۳	۱۵–۱۰

زمانی که تعداد داده‌ها و همچنین پراکندگی آن‌ها زیاد باشد از جدول طبقاتی استفاده می‌کنیم.

استفاده از جداول طبقاتی، باعث ایجاد خطایی در محاسبه شاخص های آماری خواهد شد که به آن خطای طبقه بندی گفته می شود.

(خطای طبقه‌بندی عبارت است از اختلاف بین شاخص آماری محاسبه‌شده در استفاده از توزیع فراوانی طبقه‌بندی‌شده و توزیع فراوانی طبقه‌بندی‌نشده)

1 - Error of grouping

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

به عنوان مثال، در جداول زیر که برای طبقه بندی داده های ۰ تا ۱۰۰ استفاده شده اند، جدول B مقدار خطای بیشتری دارد.

A
۱۰-۰
۲۰-۱۰
.
.
.
۱۰۰-۹۰

خطای بیشتر B
۲۵-۰
۵۰-۲۵
.
.
.

معرفی شاخص‌های توزیع فراوانی[1]

دامنه تغییرات : $R = Max - Min + 1$

به خاطر حدود واقعی

۶،۹،۱۰،۱۵،۲۶

$R = ۲۶/۵ - ۵/۵ = ۲۱$

یا $R = ۲۶ - ۶ + ۱ = ۲۱$

طول طبقات (فاصلهٔ طبقاتی) : $I = x_H - x_L + 1$

x_H: بزرگ‌ترین دادهٔ طبقه

x_L: کوچک‌ترین دادهٔ طبقه

۱: به خاطر حدود واقعی

x	f
۸ - ۱۳	
۱۴-۱۹	

$i = ۱۳ - ۸ + ۱ = ۶$

۶

تعداد طبقات $K = \frac{R}{I}$

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

به عنوان مثال، اگر دامنه تغییرات برابر ۱۰۰ و طول طبقات مساوی ۵ باشد، تعداد ۲۰ طبقه خواهیم داشت.

1 - Frequency distribution

نحوهٔ ساختن توزیع فراوانی طبقه‌بندی‌شده

اولین مرحله: محاسبه دامنهٔ تغییرات

دومین مرحله: تعیین تعداد طبقات

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

$K = ۱ + ۳/۳ \log (n)$

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

($\log^{۶۰}$ در پایه ۱۰ مساوی ۱/۷۷۸۲ است)

$K = ۱ + ۳/۳ \text{ Logn}$

$K = ۱ + ۳/۳ \log^{۶۰}$

$K = ۱ + ۳/۳ (۱/۷۷۸۲)$

$K = ۶/۸۶۸۰۶ \approx ۷$

سومین مرحله: تعیین طول طبقات $I = \frac{R}{K}$

نمونه‌ای از یک جدول توزیع فراوانی:

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	10 (=n)	۰/۳	۳۰٪	۱	۱۰۰٪

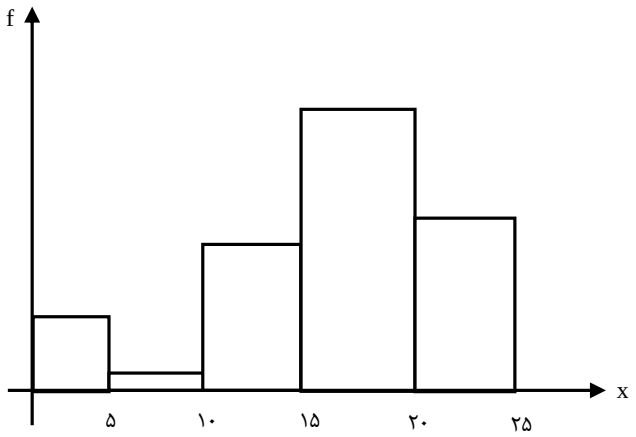
$\Sigma f = n$ $\quad \sum \frac{f}{n} = 1$ $\quad \Sigma \%f = \%100$

نکتهٔ 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 پایین‌به‌بالای جدول محاسبه خواهد شد؛ درواقع فراوانی تراکمی همواره از سمت عدد یا طبقه کوچک شروع می‌شود.

« **نمودارها**» ← نمایش گرافیکی داده‌ها (ساده‌سازی ارائه‌ی داده ها)

۱. نمودار هیستوگرام[1]

X	f
۵–۰	۴
۵–۱۰	۱
۱۰–۱۵	۷
۱۵–۲۰	۱۱
۲۰–۲۵	۸

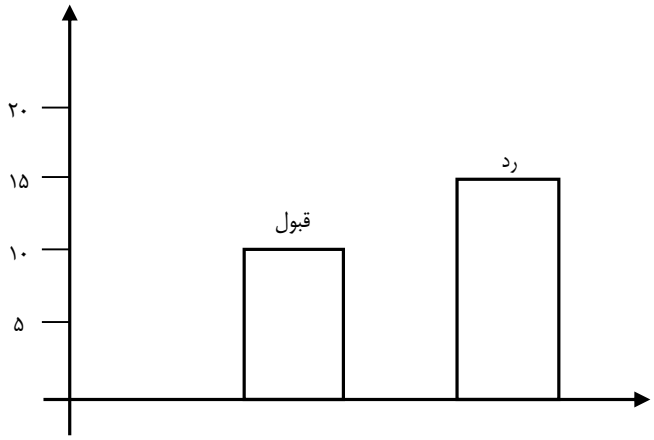


این نمودار در مقیاس‌های فاصله‌ای و نسبی (کمّی و پیوسته) استفاده شده و کاربرد مهم آن در آمار، بررسی تقارن توزیع است.

۲. نمودار ستونی[2] (میله‌ای)

جدول ۱ـ۲ توزیع فراوانی نمره‌های ۲۵ نفر در یک آزمون فیزیک

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

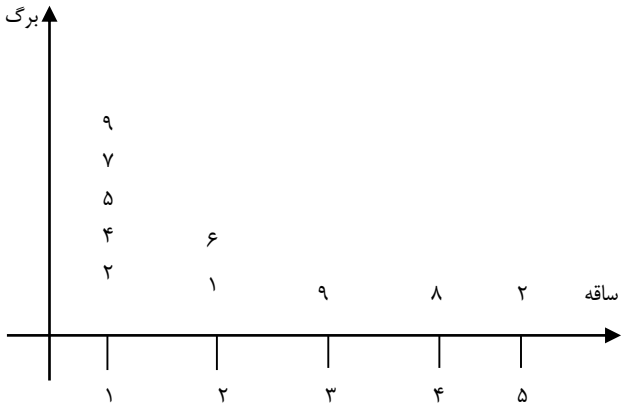


کاربرد نمودار ستونی همانند هیستوگرام بوده (بررسی تقارن توزیع) با این تفاوت که دارای مقیاس اسمی و گسسته است.
(حتی ترتیب طبقات در این نمودار اهمیتی ندارد)

1 - histogram
2 - bargraph

۳. نمودار ساقه و برگ (شاخه و برگ)[1]

۱۷ ۳۹ ۱۲ ۵۲ ۲۶
۱۵ ۲۱ ۱۹ ۴۸ ۱۴



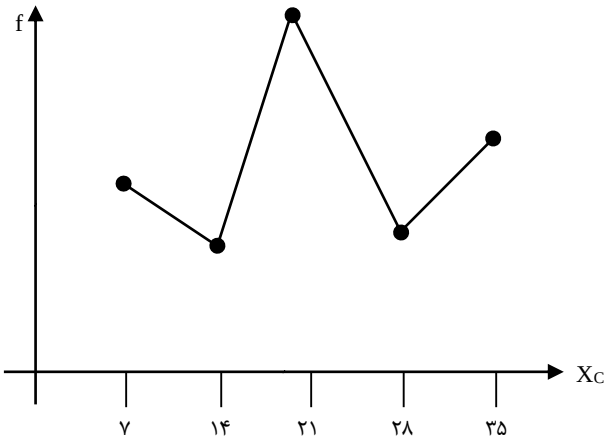
از نظر مقیاس و کاربرد، این نمودار کاملا همانند هیستوگرام است (کاربرد: بررسی تقارن- سطح سنجش: فاصله ای یا نسبی)، با این تفاوت که اگر تعداد داده‌ها کم باشد اطلاعات جزئی‌تری نسبت به هیستوگرام در اختیار قرار می‌دهد.

٤. نمودار چندضلعی[2]

نقطهٔ میانی طبقه $X_C = \frac{x_L + x_H}{۲}$

(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

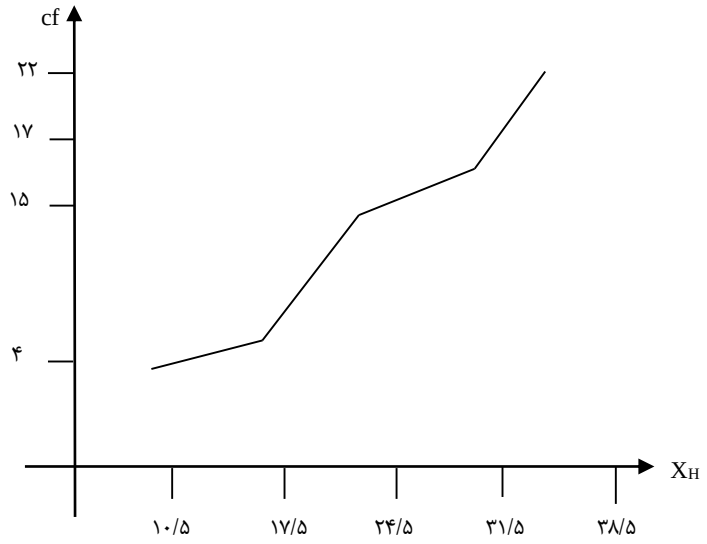


این نمودار نیز برای داده‌های کمّی و پیوسته استفاده می‌شود و پرکاربردترین نمودار به شمار می‌رود، زیرا جهت مقایسه دو یا چند توزیع از آن استفاده می‌شود.

1 - Stem and leaf plot
2 - polygon

۵. نمودار چندضلعی تراکمی (اجایو)[1]

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

٦. نمودار دایره‌ای[2]

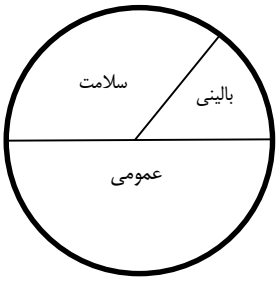
x	f
بالینی	۲۰
عمومی	۶۰
سلامت	۴۰

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

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

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

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



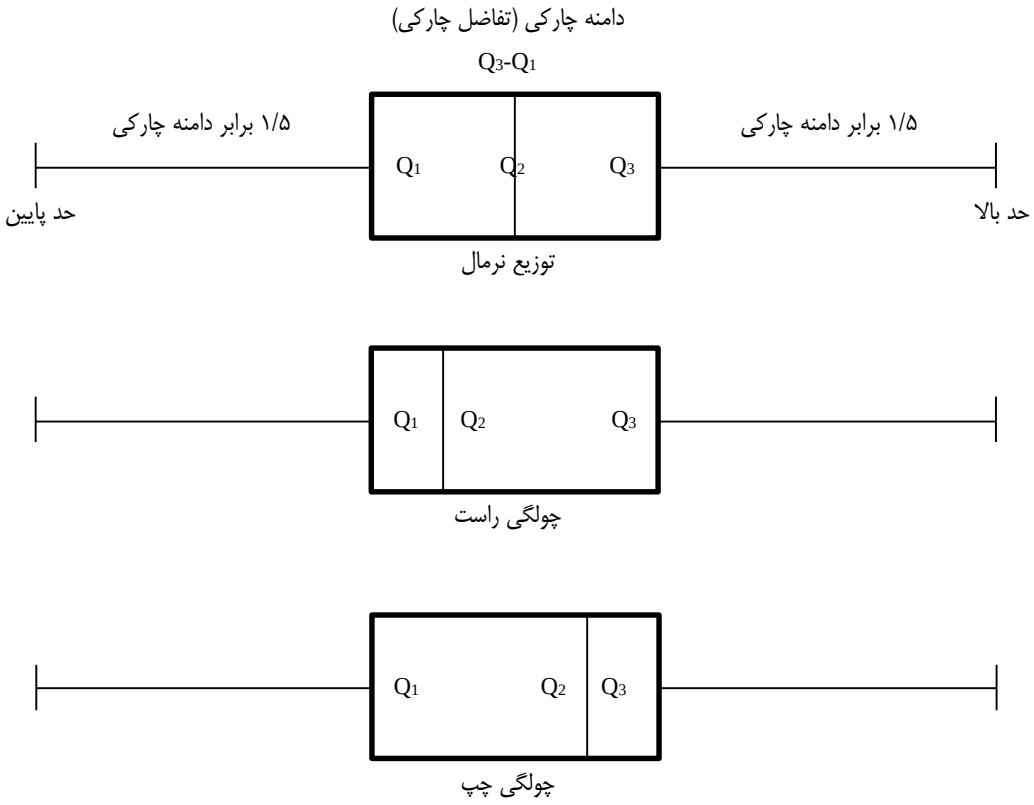
نمودار دایره‌ای دارای مقیاس اسمی - ترتیبی است و جهت نمایش نسبت جزء به کل استفاده می‌شود.

1 - cumulative (Ogive)
2 - Pie chart

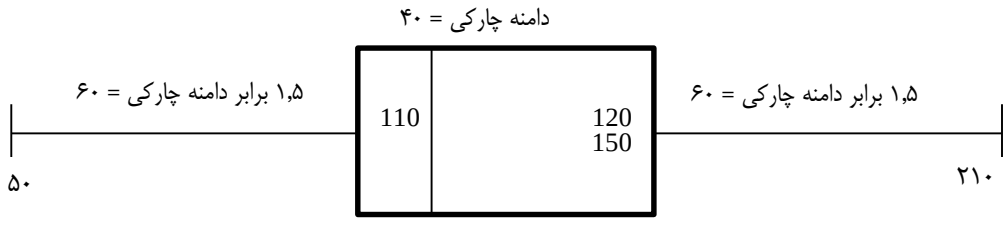
۷. نمودار جعبه‌ای[1]

این نمودار کمّی و پیوسته بوده و مهم‌ترین کاربرد آن، یافتن داده پرت است.

به‌طوری‌که ۱/۵ برابر دامنه چارکی محاسبه و از مقدار چارک اول کاسته و به مقدار چارک سوم افزوده می‌شود. دامنهٔ به‌دست‌آمده، دامنهٔ نرمال بوده و اعداد بیرون از این دامنه، دادهٔ پرت به شمار می‌روند:



– به عنوان مثال اگر در توزیع داده های یک پژوهش، مقدار چارک اول برابر ۱۱۰، چارک دوم ۱۲۰ و چارک سوم ۱۵۰ باشد، خواهیم داشت:



بدین ترتیب در این پژوهش، داده های مابین ۵۰ و ۲۱۰ داده پرت نبوده اما چنانچه بیرون از این بازه، عدد یا داده ای داشته باشیم، به عنوان داده پرت شناسایی و معرفی خواهد شد.

ضمنا از آنجایی که مقدار چارک دوم به چارک اول نزدیکتر ا ست، می توان نتیجه گرفت که بی‌شتر تراکم توزیع سمت چپ بوده و با این حساب، توزیع دارای چولگی راست یا + است.

1 - Box plot

فصـــــــــل دوم

شاخص‌های مرکزی

۱. مد[1] یا نما M_o

مد شاخصی مبنی بر تکرار است و به طبقه‌ای اطلاق می‌شود که دارای بیشترین فراوانی باشد. این شاخص ضعیف‌ترین و بی‌ثبات‌ترین 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.

مثال : 23 , 12 , 7 , 10 , 20 , 7 , 3 , 15 , 18 , 9 , 7 , 25 , 8 , 13 , 21 , 11 , 24 : $M_o = 7$

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

طرز محاسبهٔ مُد در داده‌های طبقه‌بندی‌شده (پیوسته):

برای محاسبهٔ مُد مراحل زیر را به ترتیب انجام می‌دهیم:

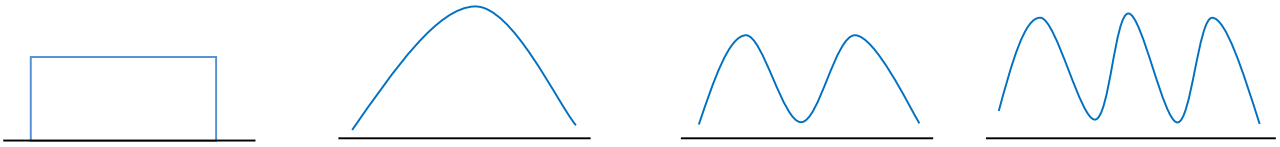
۱ـ ابتدا طبقه‌ای که بیشترین فراوانی مطلق را دارد انتخاب می‌کنیم.

۲ـ از فرمول مقابل مُد را محاسبه می‌کنیم:

حدپایین‌طبقه ← $MO = L + \frac{f_a}{f_a + f_b} \times i$ → فراوانی مطلق طبقه بالایی نما

فراوانی مطلق طبقه پایینی نما

نکته: توزیع‌های آماری مورد توجه در این کتاب، توزیع‌های تک نمایی هستند؛ اما لازم به ذکر است که توزیع فراوانی داده‌ها می‌تواند دو نمایی یا چند نمایی باشد:



توزیع بدون نما | توزیع تک‌نمایی | توزیع دونمایی | توزیع چندنمایی

۲. میانه[2] M_d

میانه نقطهٔ وسط یا پنجاه‌درصدی داده‌ها است.

این شاخص مبتنی بر ترتیب و دارای مقیاس ترتیبی ا ست؛ پس در مقیاس‌هایی می‌توان از آن ا ستفاده کرد که ویژگی ترتیب را دا شته با شند؛ یعنی مقیاس‌های ترتیبی، فاصله‌ای و نسبی.

میانه برخلاف میانگین شاخصی کیفی است که تحت تأثیر ارزش عددی یا حجم داده‌ها قرار نمی‌گیرد. لذا میانه به داده‌های 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.

1 - mode
2 - median

«محاسبه میانه»

۱. وقتی تعداد داده‌ها فرد باشد. ۱۸، ۱۵، ۱۰، ۹، ۵

۲. وقتی تعداد داده‌ها زوج باشد. ۲۰، ۱۷، ۱۴، ۱۰، ۸، ۵

از این دو میانگین می‌گیریم (۱۲) ←

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

Md

۲۰، ۱۵، ۱۲، ۸، ۸، ۸، ۷، ۵

مکان میانه: گام اول $\frac{n}{۲} = \frac{۸}{۲} = ۴$

Md

$\frac{۱}{۳}$ $\frac{۱}{۳}$ $\frac{۱}{۳}$

گام دوم :

۷/۵ ۸/۵

$$Md = 7/5 + \frac{2}{3} = 8/16$$

فرمول محاسبه میانه : $\mathbf{Md}$ = حدپایین عدد تکراری + $\frac{\text{تعداد تکرارهای قبل از میانه}}{\text{تعداد کل تکرارها}}$

x	f
۱۵	۱
۱۹	۳
۲۶	۴
۳۲	۲

Md

۱۵ ۱۹ ۱۹ ۱۹ ۲۶ ۲۶ ۲۶ ۲۶ ۳۲ ۳۲

مکان میانه : $\frac{۱۰}{۲} = ۵$

$Md = ۲۵/۵ + \frac{۱}{۴} = ۲۵/۷۵$

Md

۲۰ ، ۱۸ ، ۱۸، ۱۸، ۱۸، ۱۸ ، ۱۵ ، ۱۲

مکان میانه : $\frac{n}{۲} = \frac{۷}{۲} = ۳/۵$

$Md = ۱۷/۵ + \frac{۱/۵}{۴} = 17/87$

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

$$Md = L + \frac{\frac{n}{۲} - cf}{f}(i)$$

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

cf = فراوانی تراکمی طبقهٔ قبل از طبقه میانه

f = فراوانی مطلق طبقه‌ای که میانه در آن قرار دارد

i = فاصلهٔ طبقات

n = تعداد نمره‌ها

به‌طور 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
۲۷_۲۹	۶	۱۲
۲۴_۲۶	۳	۶
۲۱_۲۳	۲	۳
۱۸_۲۰	۱	۱

محاسبهٔ میانهٔ برای توزیع فراوانی بالا، شامل پنج مرحله به شرح زیر است:

۱. تقسیم تعداد کل نمره‌ها بر ۲ ۵۰÷ ۲ = ۲۵

۲. مراجعه به ستون فراوانی تراکمی و پیدا کردن اولین طبقه‌ای که فراوانی تراکمی آن مساوی یا بزرگ‌تر از $\frac{N}{۲}$ 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 = ۳۲/۵

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.

$$Md = ۳۲/۵ + \frac{\frac{۵۰}{۲} - ۱۹}{۸}(۳) = ۳۴/۷۵$$

۳. میانگین[1] [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.

$X =$ ۲ ۳ ۵ ۱۴ $\quad \bar{X} = ۶ \quad X - \bar{X}$: −۴ −۳ −۱ ۸

−۸ +۸

همواره و در هر توزیعی: $\sum(x - \bar{x}) = 0$

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

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

هرعدد دلخواهی : a

محاسبه میانگین:

برای داده های خام : $\bar{x} = \frac{\sum x}{n}$

در جدول غیر طبقاتی : $\bar{x} = \frac{\sum f x}{n}$

X	f
۱۵	۳
۱۶	۴
۱۰	۲

برای جدول طبقاتی : $\bar{x} = \frac{\sum f x_c}{n}$

X	f	X_C
۱۵ _۲۳	۳	۱۹
۲۴_۳۲	۱	۲۸
۳۳_۴۱	۴	۳۷

1 - mean

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

$\overline{X}=۱۵$ $\overline{X}=۱۰$

$n=۱۰$ $n=۴۰$

میانگین وزنی (میانگین کل، میانگین مرکب، میانگین میانگین‌ها) $\overline{x}=\frac{n_۱\overline{x}_۱+n_۲\overline{x}_۲+\dots}{n_۱+n_۲+\dots}$

$$\text{میانگین کل}=\frac{۱۰\times ۴۰+۱۵\times ۱۰}{۵۰}=\frac{۵۵۰}{۵۰}=۱۱$$

فرمول تجربی پیرسون (رابطهٔ میان شاخص‌های مرکزی) :

$$m_۰=۳m_d-۲\overline{x}$$

هرگاه بخواهیم به‌کمک دو شاخص مرکزی، شاخص سوم را به‌دست بیاوریم، می‌توان با تقریب نسبتا مناسبی از فرمول تجربی پیرسون استفاده کرد.

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

$$۵=۳Md-2\times 20 \quad : \quad Md=۱۵$$

میانگین هندسی[1]

برای محاسبه میانگین درصدها، نسبت‌ها یا نرخ‌ها استفاده می‌شود.

$$G=\sqrt[n]{x_۱x_۲\dots x_n}$$

مثال: میانگین هندسی اعداد ۱۲۵ ۲۷ ۸

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

میانگین هارمونیک[2] یا همساز

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

$$HM \text{ یا } H=\left(\frac{n}{\frac{۱}{x_۱}+\frac{۱}{x_۲}+\dots+\frac{۱}{x_n}}\right)$$

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

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

فرمول محاسبه میانگین همساز از طریق میانگین‌های هندسی و حسابی.

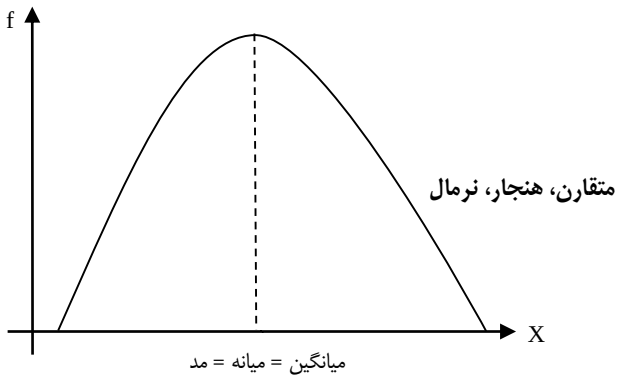
$$HM=\frac{G^۲}{\overline{X}}$$

1 - Geometric mean
2 - Harmonic mean

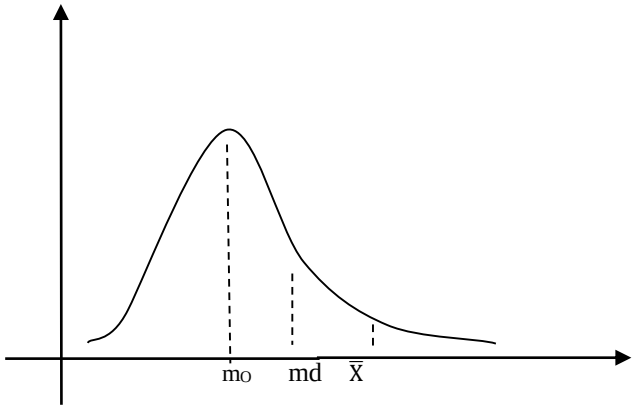
چولگی[1] (کجی)

شاخص چولگی تحت تأثیر داده پرت است و جهت بررسی تقارن توزیع 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



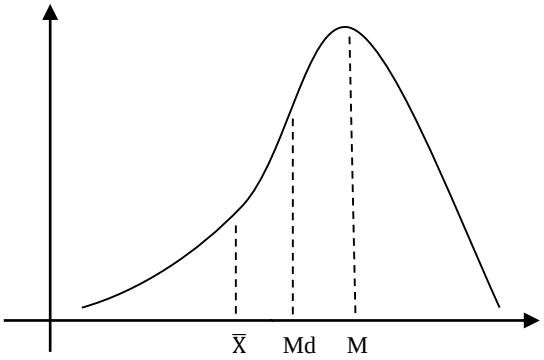
چولگی راست یا مثبت:

بااینکه اکثریت نمره کم دارند، به خاطر چند تا نمره بزرگ میانگین بزرگ‌تر از نما و میانه شده است.

دادهٔ پرت بزرگ ⟵ میانگین را بزرگ می‌کند ⟵ چولگی راست

مثال برای چولگی راست: توزیع درآمد در تهران

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



چولگی چپ یا منفی

دادهٔ پرت کوچک ⟵ میانگین را کوچک می‌کند ⟵ چولگی چپ

1 - skewness

محاسبه چولگی

$$sk = \frac{\bar{x} - m_{\cdot}}{s} \qquad |sk| \leq ۱ \text{ نرمال}$$

اگر شدت چولگی محاسبه شده توسط این فرمول کمتر از عدد یک (در بعضی منابع عدد دو) باشد، مقدار چولگی خفیف بوده و می‌توان توزیع را نرمال درنظر گرفت.

چندک‌ها شاخص‌های وضعی و هم‌خانوادهٔ میانه هستند، به‌طوری‌که 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.

۷ ، ۹ ، ۹ | ۹ ، ۱۲ ، ۱۸ | ۲۰ ، ۲۵ ، ۲۵ | ۲۵ ، ۲۵ ، ۲۵

$Q_1 = 8/5 + \frac{2}{3}$ $\quad Q_۲ = ۱۹$ $\quad Q_۳ = ۲۴/۵ + \frac{۲}{۵}$

فرمول محاسبه چندک‌ها:

همان فرمول محاسبه میانه است، با این تفاوت که به‌جای n تقسیم بر ۲، باید نسبت مورد نظر را جای‌گذاری کرد. برای مثال برای محاسبهٔ 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 به‌جای $\frac{n}{۴}$ مقدار $\frac{۳n}{۴}$ استفاده کنیم.

یا برای محاسبه دهک هفتم: $\frac{۷n}{۱۰}$

برای محاسبه نقطه ۸۴ درصدی: $\frac{۸۴n}{۱۰۰}$

1 - quartile
2 - decile
3 - percentile

فصـــــــــل سوم

شاخص‌های پراکنـدگی

شاخص‌های پراکندگی[1] (فاصله‌ای)

۱. دامنه تغییرات[2]

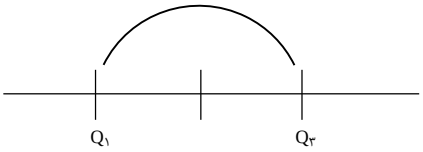
$$R = max - min + ۱$$

پراکنده‌تر : R = ۱۸ : کلاس A ۱۹ و ..... و ۳ و ۲

همگن‌تر : R = ۵ : کلاس B ۱۸ و ..... و ۱۴ و ۱۴

نه‌تنها دامنهٔ تغییرات بلکه تمامی شاخص‌های پراکندگی به دنبال فاصلهٔ میان داده‌ها و دارای مقیاس فاصله‌ای هستند. دامنهٔ تغییرات ضعیف‌ترین و بی‌ثبات‌ترین شاخص پراکندگی به شمار می‌رود، در محاسبهٔ 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، دامنهٔ 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[3]



$$Q = \frac{Q_۳ - Q_۱}{۲}$$ (چارک متوسط)

انحراف چارکی نشـان می‌دهد داده‌ها به‌طور متوسـط چقدر از میانه یا مرکز توزیع انحراف دارند. درواقع انحراف چارکی میزان پراکندگی را در ۵۰٪ وسـط توزیع بررسـی کرده لذا به داده‌های پرت کوچک یا بزرگ حسـاس نیسـت. پس می‌توان گفت زمانی که توزیع داده‌ها چولگی قابل‌ملاحظه‌ای داشـته باشـد، بهترین و مناسب‌ترین شاخص پراکندگی انحراف چارکی خواهد بود.

۳. انحراف متوسط[4]

$$AD = \frac{\sum|x - \bar{x}|}{n}$$ یا MD

انحراف متوسط عبارت است از متوسط قدر مطلق انحراف داده‌ها از میانگین. این شاخص نشان می‌دهد که داده‌ها به‌طور متوسط چقدر از میانگین انحراف دارند. انحراف متو سط یک ایراد ا سا سی دارد و آن اینکه به کمک قراردادی به نام قدر مطلق علائم جبری را از بین برده و در عملیات ریا ضی و آماری نمی‌توان از آن استفاده کرد.

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

از آنجایی که میانگین مرکز ثقل بوده و مقدار انحرافات منفی و مثبت از آن به یک اندازه است، خواهیم داشت:

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

1 - variability
2 - range
3 - Quartile diviation
4 - Mean diviation

٤. واریانس[1]

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

واریانس عبارت است از متوسط مجذور انحراف داده‌ها از میانگین.

دلیل نام‌گذاری واریانس: S^2 ⟵ **SS 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.

مثلاً $m \longrightarrow m^2$

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) به طور تجربی 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 اُریب را از بین ببریم لازم است در مخرج فرمول از $n-1$ استفاده کنیم:

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

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

نمونه است و می‌خواهیم تعمیم دهیم.

واریانس نمونه

$X = ۲ \quad ۳ \quad ۶ \quad ۸ \quad ۱۶$

$\bar{X} = \frac{۳۵}{۵} = ۷$

$(x-\bar{x})$	$(x-\bar{x})^2$	$\sum(x-\bar{x})^2 = ۱۲۴$
$۲-۷=-۵$	۲۵	
$۳-۷=-۴$	۱۶	
$۶-۷=-۱$	۱	
$۸-۷=۱$	۱	
$۱۶-۷=۹$	۸۱	

-۱۰ و $+۱۰$

واریانس

- نمونه (برآورد، تعمیم): $S^2 = \frac{\sum(x-\bar{x})^2}{n-1} = \frac{124}{5-1} = ۳۱$
- جامعه (واریانس واقعی): $S^2 = \frac{\sum(x-\bar{x})^2}{n} = \frac{۱۲۴}{۵} = ۲۴/۸$

محاسبه واریانس به کمک اعداد خام:

$$S^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{369-\frac{1225}{5}}{4} = \frac{124}{4} = 31$$

1 - variance

محاسبه واریانس مرکب

همان‌طور که در فصل قبلی به محاسبه میانگین کل (میانگین مرکب) برای دو یا چند میانگین (دو یا چند گروه) اشاره شد، زمانی‌که محقق علاقمند باشد واریانس مرکب دو یا چند گروه را محاسبه کند، می‌تواند از فرمول زیر که به فرمول مک نیمار[۱] معروف است، استفاده کند:

$$S_T^2 = \frac{n_A(\bar{X}_A^2+S_A^2)+ n_B(\bar{X}_B^2+S_B^2)}{n_A+n_B} - \bar{X}_T^2$$

S_T^2 : واریانس مرکب

$\bar{X}_T^2$: میانگین مرکب

نکته: در صورتی‌که بیشتر از دو گروه داشته باشیم، عناصر گروه‌های دیگر به صورت و مخرج فرمول فوق اضافه می شوند. ضمنا اگر قصد محاسبه انحراف استاندارد مرکب را داشته باشیم، از تمام فرمول مک نیمار، جذر یا رادیکال گرفته و بدین‌ترتیب انحراف استاندارد مرکب محاسبه خواهد شد.

۵. انحراف استاندارد[۲] (انحراف‌معیار) $S = \sqrt{S^2}$

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.

سوال: در توزیعی با کجی SK = – ۰/۷۳ چنانچه بهترین شاخص گرایش به مرکز، میانگین انتخاب شده باشد. بهترین شاخص پراکندگی کدام خواهد بود؟

پاسخ: S : زمانی که میانگین بهترین شاخص مرکزی است، توزیع نرمال بوده و بهترین شاخص پراکندگی انحراف معیار خواهد بود:

	مرکزی	پراکندگی
نرمال	$\overline{X}$	S
غیرنرمال	Md	Q

مقدار SK یا چولگی نیز در این سوال کمتر از یک بوده و نشان از نرمال بودن توزیع دارد.

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

(مقدار انحراف چارکی از سایر شاخص های پراکندگی کمتر بوده و مقدار انحراف معیار (S) نیز همواره کمی بیشتر از انحراف متوسط (MD) است)

در واقع، در یک توزیع طبیعی، تقریباً ۵۰ درصد داده‌ها در فاصله‌ی ±۱Q ، ۵۸ درصد در فاصله‌ی ±۱MD و ۶۸ درصد داده‌ها در فاصله‌های ±۱S از میانگین قرار دارند.

تصحیح شپرد[۳] $S = \sqrt{S^2 - \frac{i^2}{12}}$

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 مجذور 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 < ۱۲

۱ – Mc Nemar

2 - Standard deviation

3 - Sheppard's correction

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

$S = \sqrt{۲۸ - \frac{۳۶}{۱۲}} = \sqrt{۲۵} = ۵$

$i = ۶$

$S^۲ = ۲۸$

$۹ < ۱۲$

اگر $K > ۱۲$ بود دیگر نیازی به تصحیح نداشت: $S = \sqrt{۲۸}$

ضریب تغییرات[1] (ضریب پراکندگی) $V = \frac{S}{\bar{X}}$ یا c.v

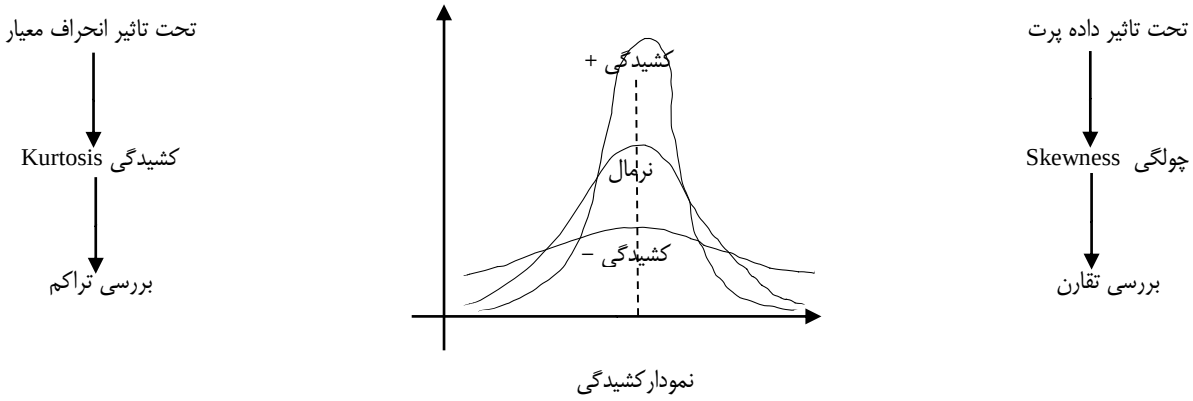
برای مقایسه پراکندگی یک ویژگی در گروه‌های مختلف، می‌توان از S یا S^2 استفاده کرد؛ اما برای مقایسه پراکندگی دو یا چند ویژگی مختلف با واحدهای متفاوت می‌بایست از شاخصی به نام ضریب تغییرات که مستقل از واحد اندازه‌گیری است استفاده کرد.

برای مثال:

دانش‌آموزان یک کلاس

- وزن
  - $S = ۲۰$
  - $\bar{X} = ۵۰$
  - $V = \frac{s}{\bar{x}} = \frac{۲۰}{۵۰} =$ ٪۴۰ یا ۰/۴ پراکنده‌تر
- قد
  - $S = ۳۰$
  - $\bar{X} = ۱۰۰$
  - $V = \frac{s}{\bar{x}} = \frac{۳۰}{۱۰۰} =$ ٪۳۰ یا ۰/۳ همگن‌تر

کشیدگی[2]



نمودار کشیدگی

کشیدگی، تحت تأثیر مقدار انحراف استاندارد بوده و جهت بررسی تراکم توزیع از آن استفاده می شود؛ به‌طوری‌که مقدار کشیدگی، با میزان تراکم توزیع رابطه مستقیم اما با میزان پراکندگی (S) رابطه عکس دارد. به طوری که در کشیدگی منفی یا پخ (هموار)، مقدار انحراف معیار زیاد بوده (تراکم کم) در حالی که در کشیدگی مثبت یا قله ای، مقدار انحراف معیار کم اما تراکم زیادی در توزیع داده ها وجود دارد.

1 - Coefficient of variation
2 - kurtosis

نکته: توزیع نرمال به توزیعی گفته می‌شود که نه چولگی شدید و نه کشیدگی شدید داشته باشد.

گشتاور[1]:

گشتاور مرتبه اول $\frac{\sum(x-\bar{x})}{n} = 0 : m_1$

گشتاور مرتبه دوم $\frac{\sum(x-\bar{x})^2}{n} = \frac{n-1}{n}S^2 : m_2$

گشتاور مرتبه سوم $\frac{\sum(x-\bar{x})^3}{n} : m_3$

و به همین ترتیب تا مرتبه n ام.

فرمول های چولگی و کشیدگی بر حسب گشتاور: چولگی: $Sk = \frac{m_3}{S^3}$ کشیدگی: $Ku = \frac{m_4}{S^4} - 3$

همان‌طور که ملاحظه می‌شود، چولگی (کجی)، تحت تأثیر گشتاور مرتبه سوم بوده و کشیدگی، از گشتاور مرتبه چهارم تأثیر می‌پذیرد.

تأثیر اعمال ریاضی بر شاخص‌های مرکزی و پراکندگی:

به‌طورکلی شاخص‌های مرکزی به تمامی اعمال ریاضی حساس بوده درحالی‌که شاخص‌های پراکندگی فقط به ضرب و تقسیم حساس هستند.

چنانچه تمامی داده‌های یک توزیع را در عدد ثابتی ضرب یا تقسیم کنیم، شاخص‌های پراکندگی نیز در همان عدد ضرب یا تقسیم خواهند شد به‌جز واریانس که در مجذور آن عدد ضرب یا تقسیم می‌شود.

تست: اگر به بزرگ‌ترین داده یک توزیع ۱۰ نمره اضافه شود، کدام‌یک از شاخص‌های زیر بدون تغییر باقی می‌ماند؟

۱) میانگین ۲) انحراف استاندارد ۳*) میانه ۴) انحراف متوسط

سؤال: تمامی داده‌های 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؟

$2X+5$

$2\overline{X}+5=25 \qquad 2\overline{X}=20 \qquad \overline{X}=10$

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

سؤال: اگر مقدار انحراف معیار x_n و ... و x_2 و x_1 برابر ۴ باشد، مقدار واریانس $10+5x_n$ و ... و $10+5x_1$ چقدر است؟

$S = 4$

$\downarrow \times 5$

$S_x = 4 \times 5 = 20 \qquad S = 20 : S^2 = 400$

مثال: فراوانی تمامی داده‌ها را در عدد ۳ ضرب کرده‌ایم، کدام تغییر دربارهٔ میانگین اتفاق می‌افتد؟

۱) افزایش ۲) کاهش ۳) تغییر نمی‌کند ✓ ۴) تغییر می‌کند

1 - moment

نکته مهم: درصورتی‌که **فراوانی** تمام داده‌ها به شکل یکسانی تغییر کند **هیچ شاخص آماری تغییر نمی‌کند.**

سؤال: اگر داده‌های یک توزیع، همگی با عدد ثابت مثبتی جمع شوند، مقدار ضریب تغییرات چه تغییری خواهد کرد؟

پاسخ: کاهش می‌یابد.

ازآنجاکه تنها شاخص‌های مرکزی به جمع و تفریق حساس هستند، پس در صورت اضافه شدن یک عدد به تمامی داده‌ها، در فرمول ضریب تغییرات مقدار مخرج (میانگین) افزایش پیدا می‌کند، درحالی‌که انحراف معیار بدون تغییر باقی می‌ماند، به همین دلیل مقدار ضریب تغییرات کاهش پیدا خواهد کرد.