بسم الله الرحمن الرحیم

مؤسسه آموزش عالی آزاد ماهان با افتخار تقدیم می‌کند

کاربرد آمار و روش تحقیق

مجموعه ارتباطات اجتماعی

از سری کتاب‌های کمک آموزشی کارشناسی ارشد

حسین صنعتی
پریسا حاج‌کرمی

موسسه آموزش عالی آزاد
ماهان
www.mahan.ac.ir

سرشناسه	صنعتی، حسین
عنوان	کاربرد آمار و روش تحقیق / حسین صنعتی-پریسا حاج کرمی
مشخصات نشر	تهران، مشاوران صعود ماهان، ۱۴۰۴.
مشخصات ظاهری	۴۰۶ صفحه رحلی
فروست	سری کتاب‌های کمک آموزشی کارشناسی ارشد
شابک	۹۷۸-۶۰۰-۳۸۹-۲۴۶-۰
وضعیت فهرست‌نویسی	فیپای مختصر
یادداشت	این مدرک در آدرس http://opac.nali.ir قابل دسترسی است
شناسه افزوده	حاج کریمی، پریسا
شماره کتاب‌شناسی ملی	۳۲۰۰۳۴۱

ماهان
مشاوران صعود

مجموعه ارتباطات اجتماعی

کاربرد آمار و روش تحقیق

ناشر	مشاوران صعود ماهان
مدیر مسئول	هادی سیاری - مجید سیاری
مدیر برنامه‌ریزی و تولید	محسن صداقت
به قلم	حسین صنعتی - پریسا حاج‌کرمی
ویراستار ادبی	زهرا سروری
گرافیست	حامد شاملو
صفحه‌آرا	فاطمه طاهر
طراح جلد	سمیرا خانزاد
نوبت و تاریخ چاپ	اول / ۱۴۰۴
شمارگان	۲۰۰۰ نسخه
قیمت	۶/۶۲۰/۰۰۰ ریال
شابک	۹۷۸-۶۰۰-۳۸۹-۲۴۶-۰

نشانی: تهران، خیابان ولی‌عصر- بالاتر از تقاطع مطهری- جنب بانک ملی پلاک ۲۰۵۰
شماره تماس:
۸۸۱۰۰۱۱۳-۴

سخنِ ناشر

«ن والقلم وما یسطرون»

کلمه نزد خدا بود و خدا آن را با قلم بر ما نازل کرد.

به پاس تشکر از چنین موهبت الهی، مؤسسه ماهان درصدد برآمده است تا در راستای انتقال دانش و مفاهیم با کمک اساتید مجرب و مجموعه کتب آموزشی خود برای شما داوطلبان ادامه تحصیل در مقطع کارشناسی ارشد، گام مؤثری بردارد. امید است تلاش‌های خدمتگزاران شما در این مؤسسه پایه‌گذار گام‌های بلند فردای شما باشد.

مجموعه کتاب‌های کمک آموزشی ماهان به‌منظور استفاده داوطلبان کنکور کارشناسی ارشد سراسری و آزاد تألیف شده‌است. در این کتاب‌ها سعی کرده‌ایم با بهره‌گیری از تجربه اساتید بزرگ و کتب معتبر داوطلبان را از مطالعه کتاب‌های متعدد در هر درس بی‌نیاز کنیم.

دیگر تألیفات ماهان برای دانشجویان به‌صورت ذیل است:

مجموعه کتاب‌های **۸** آزمون: شامل **۵** مرحله کنکور کارشناسی ارشد **۵** سال اخیر به همراه **۳** مرحله آزمون تألیفی ماهان همراه با پاسخ تشریحی است که برای آشنایی با نمونه سؤالات کنکور طراحی شده است. این مجموعه کتاب‌ها با توجه به تحلیل **۳** ساله اخیر کنکور و بودجه‌بندی مباحث در هریک از دروس، اطلاعات مناسبی برای برنامه‌ریزی درسی در اختیار دانشجو قرار می‌دهد.

مجموعه کتاب‌های کوچک: شامل کلیه نکات کاربردی در گرایش‌های مختلف کنکور کارشناسی ارشد است که برای دانشجویان به‌منظور جمع‌بندی مباحث در **۲** ماهه آخر قبل از کنکور مفید است.

بدین‌وسیله از مجموعه اساتید، مؤلفان و همکاران محترم خانواده بزرگ ماهان که در تولید و بروزرسانی تألیفات ماهان نقش مؤثری داشته‌اند، صمیمانه تقدیر و تشکر می‌نماییم.

دانشجویان عزیز و اساتید محترم می‌توانند هرگونه انتقاد و پیشنهاد در خصوص تألیفات ماهان را از طریق سایت ماهان به آدرس mahan.ac.ir با ما در میان بگذارند.

مؤسسه آموزش عالی آزاد ماهان

مقدمهٔ مؤلف

دوست گرامی!
با عرض سلام و احترام
سعی ما در این کتاب بر این بوده است که مهم‌ترین مطالب را از منابع آزمونی درس کاربرد آمار و روش تحقیق در علوم ارتباطات اجتماعی فراهم آوریم و با زبانی روان تشریح کرده و تقدیم شما نماییم.

هریک از بخش‌ها و یا فصل‌های این کتاب دارای تست‌های تمرینی است که هدف از طرح آنها کمک به خواننده در افزایش تسلط بر مطالب مربوطه از زوایای مختلف است. در آخر هرکدام از بخش‌ها نیز سوالات کنکوری سراسری و آزادِ مربوط به همان بخش درج شده و به صورت مشروح پاسخ داده شده‌اند.
در تدوین کتاب حاضر، آخرین تغییرات در منابع طرح سوالات آزمونی لحاظ شده و مطالب آزمونی مفید موجود در منابع جدید نیز درج شده‌اند. چنانچه شما عزیزان مایل به حصول اطلاع از آخرین تغییرات در منابع این درس و یا رفع سوالات و ابهامات خود در خصوص این کتاب باشید، می‌توانید از طریق پست الکترونیکی با نگارنده این سطور تماس حاصل فرمایید.
صمیمانه آرزوی موفقیت شما را داشته و انتقادات و پیشنهاداتتان برای ارتقای کیفیت این مجموعه را به دیده منت پذیراییم.

حسین صنعتی
Hosein.sanati62@gmail.com

فهرستِ مطالب

بخش اول: آمار

مهم‌ترین عناوین بخش

- مفاهیم اساسی آمار
- آماده‌سازی، سازمان‌دهی و نمایش داده‌ها (نمودارها)
- نشان دادن مرکزیت داده‌ها (پارامترهای مکان مرکزی)
- پارامترهای پراکندگی
- آمار استنباطی (برآورد و آزمون فرضیه) و ضرایب همبستگی
- تخمین آماری پارامترهای جامعه
- احتمال
- نکات آماری مهم

فصل اول

مفاهیم اساسی آمار

تعریف علم آمار

آمار علمی است که با استفاده از فنون و روش‌های علمی و ریاضی به جمع‌آوری، طبقه‌بندی و پیش‌بینی اطلاعات کمّی و کیفی و نتیجه‌گیری و تعمیم آنها در جهت هدف معین می‌پردازد.

آمار نه تنها در پژوهش‌های اجتماعی بلکه در سایر پژوهش‌های علمی نیز به مثابه وسیله است نه هدف. آمار شاخه‌ای از روش‌شناسی علمی است که با جمع‌آوری، طبقه‌بندی، توصیف و تفسیر داده‌هایی که از اجرای زمینه‌یابی‌ها و آزمایش‌ها به‌دست می‌آید، سر و کار دارد و هدف اساسی آن، توصیف و بیان استنباط‌هایی درباره ویژگی‌های عددی جوامع است.

به طور کلی آمار را می‌توان علم و عمل استخراج، بسط و توسعه دانش‌های تجربی انسانی با استفاده از روش‌های گردآوری، تنظیم، پرورش و تحلیل داده‌های تجربی دانست.

ویژگی منحصر به فرد آمار این است که نمودهای عددی بزرگ اجتماعی را در مرکز میدان دید خود می‌نهد و عمدتاً به پدیده‌هایی می‌پردازد که قابل شمارش، قابل اندازه‌گیری یا قابل کمیت‌یابی هستند.

آمار توصیفی و آمار استنباطی

همان‌طور که از تعریف علم آمار معلوم می‌شود فعالیّت آماری دو بخش را در برمی‌گیرد: الف- توصیف، ب- تبیین و برآورد (استنباط).

الف) جمع‌آوری، تلخیص، تنظیم و ارائه‌ی اطلاعات به صورت روشن و قابل درک و در صورت لزوم تعیین روابط موجود بین اطلاعات جمع‌آوری‌شده. این بخش از آمار را که بیشتر به مشخص کردن داده‌ها، تنظیم و ارائه آنها به صورت جدول‌بندی یا ترسیمی، محاسبه آماره‌ها و تعیین ارتباط بین اطلاعات می‌پردازد، آمار توصیفی می‌نامند.

آمار توصیفی برای توصیف گروهی که مورد مطالعه‌ی محقق قرارگرفته‌اند، به‌کار می‌رود. هدف در آمار توصیفی، تقلیل، توصیف، تلخیص و نمایش روابط بین متغیرها است.

در واقع هنگامی که توده‌ای از اطلاعات کمّی (مقداری) برای تفسیر گردآوری می‌شود، ابتدا لازم است آنها به صورتی که به روشنی قابل فهم و انتقال باشند سازمان بندی و خلاصه شوند که آمار توصیفی به همین منظور به‌کار برده می‌شود.

شاخص‌هایی که در آمار توصیفی به کار برده می‌شوند عبارتند از: مد، میانه، معدل و همبستگی. توجه شود که روش‌های آمار توصیفی همیشه برای تعیین و بیان ویژگی ها یا اطلاعاتی که به وسیله‌ی پژوهشگران جمع‌آوری شده‌اند به کار برده می‌شوند.

ب) چنانچه محقق قصد داشته باشد که با مطالعه بر روی نمونه‌ای منتخب از یك جامعه‌ی آماری، نتایج حاصله را به آن جامعه آماری تعمیم دهد، باید از روش‌های آمار استنباطی استفاده نماید. در بیشتر فعالیت‌های آماری جمع‌آوری، تنظیم و ارائه یافته‌ها و یا تعیین آماره‌ها و روابط بین داده‌ها کفایت نمی‌کند، بلکه لازم است بر اساس این اطلاعات جمع‌آوری و تنظیم شده، تجزیه و تحلیل‌ها و استنباط‌هایی برای تبیین و تصمیم‌گیری صورت گیرد. این بخش از آمار که بر تحلیل، تفسیر و تعمیم نتایج حاصل از تنظیم و محاسبه مقدماتی آماری تأکید دارد، آمار استنباطی خوانده می‌شود.

در آمار استنباطی محقق به چیزی خارج از گروهی از موضوعات یا افرادی که در حقیقت مورد آزمون قرار می‌گیرند نظر دارد، و یا به عبارت دیگر، ویژگی‌های نمونه آماری را به جامعه آماری تعمیم می‌دهد. در آمار استنباطی فرض بر این است که مطالعه، یک مطالعه نمونه ای است.
هدف عمده در آمار استنباطی، استنباط پارامتر از آماره است، و در واقع، در آمار استنباطی آماره‌ها را به پارامترها تعمیم می‌دهیم.

نکته ۱: در آمار توصیفی، آما ردان به توصیف خصایص نمونه‌ها می‌پردازد و در آمار استنباطی مشخصات جامعه را از روی نمونه‌ها استنباط یا برآورد می‌کند.

نکته ۲: آمار توصیفی در خدمت آمار استنباطی است.

برای درک کامل اختلاف بین آمار توصیفی و آمار استنباطی نیاز به بحث درباره برخی از مفاهیم است.

تعریف برخی از مفاهیم آماری

جامعه آماری: مجموعه‌ای از اشیاء، حوادث و افراد جامعه که در یک یا چند صفت مشترک باشند را یک جامعه آماری می‌گویند.
نمونه آماری: مجموعه‌ای از افراد هستند که از جامعه آماری انتخاب می‌شوند. نمونه آماری باید معرف جامعه آماری باشد، یعنی باید ویژگی‌ها و خصوصیات جامعه آماری را در دل خود داشته باشد.
آماره: آماره‌ها اندازه‌گیری‌هایی هستند که ویژگی‌های نمونه آماری را بیان می‌کنند و از آنها برای توصیف ویژگی‌های نمونه آماری استفاده می‌شود. معمولاً آماره‌ها با حروف لاتین نمایش داده می‌شوند.
پارامتر: پارامترها اندازه‌گیری‌هایی‌اند که ویژگی‌های یک جامعه آماری را توصیف می‌کنند و معمولاً با حروف یونانی به نمایش در می‌آیند.
پارامترها ثابت هستند اما آماره‌ها از نمونه‌ای به نمونه‌ی دیگر متفاوت می‌باشند.
نمونه‌گیری اریب و نااریب: یك نمونه‌ی تصادفی یا نااریب (Unbiased Sample) نمونه‌ای است که تمامی افراد جامعه، شانس مساوی انتخاب شدن در آن را داشته باشند. در مقابل، نمونه اریب (Biased Sample) نمونه‌ای است که بعضی از افراد جامعه، شانس بیشتری برای انتخاب شدن پیدا می‌کنند.
قضیه‌ی حد مرکزی: اگر تعداد دفعات نمونه‌گیری از یك جامعه به سمت بی نهایت میل کند، توزیع میانگین‌های نمونه به سمت نرمال میل خواهد نمود.
اولین قدم در پژوهش، تعریف جامعه و سپس انتخاب نمونه است.
نمونه: نمونه عبارت است از زیر مجموعه‌ای که از کل جامعه انتخاب می‌شود و معرف آن است.

متغیر و انواع آن

اصطلاح متغیّر به خصوصیّتی اشاره دارد که به وسیله‌ی آن اعضای یک گروه یا مجموعه از هم متمایز می‌شوند. متغیّر کمیّتی است که می‌تواند از یک فرد به یک فرد دیگر و یا از یک مشاهده به یک مشاهده دیگر، مقادیر مختلفی را اختیار کند و به بیان دقیق‌تر، نمادی است که اعداد یا ارزش‌ها بدان انتساب داده می‌شود. به مجموعه‌ای از مقادیر یك متغیر، توزیع (Distribution) آن متغیر گفته می‌شود.

انواع متغیر از حیث قابلیت عددپذیری

متغیر کیفی: متغیری است که اندازه‌های آن به صورت غیرعددی بیان می‌شود، نظیر جنسیت، رضایت شغلی، میزان مشارکت سیاسی و سطح سواد.
متغیرهای کیفی متغیرهایی هستند که از یک رشته طبقات تشکیل می‌شوند که باید کامل و مانعه‌الجمع باشند. یعنی مثلاً متغیر جنسیت که هم شامل زن و هم شامل مرد شود و یا متغیر مذهب که شامل کلیه مذاهب باشد. مثلاً اگر ما دین (مذهب) را به صورت مسلمان یا مسیحی بیاوریم در اینجا کامل نیست چون ممکن است فردی زرتشتی یا یهودی باشد. منظور از مانعه‌الجمع بودن هم یعنی اینکه مثلاً اگر بگوییم مسیحی، مسلمان، شیعه، زرتشتی، اشتباه است چون دو طبقه آن یعنی شیعه و مسلمان با هم مانعه‌الجمع نمی‌باشند و شیعه‌ها، مسلمان هستند.
این متغیرها قابل تبدیل‌شدن به عدد نیستند و اگر چه به هر طبقه‌ای، شماره‌ای داده می‌شود ولی این شماره جنبه ریاضی ندارد و به معنای کدگذاری طبقات می‌باشد. مثلاً همچنان که گفتیم، جنسیت یک متغیر کیفی است که به زن و مرد طبقه‌بندی می‌شود که ما می‌توانیم به زن کد ۱ و به مرد کد ۲ را بدهیم. در اینجا عددهای ۱و ۲ جنبه‌ی ریاضی ندارند و صرفاً جهت کدگذاری به‌کار رفته‌اند.
پس می‌توان گفت متغیرهای کیفی متغیرهایی هستند که پژوهش‌گر نمی‌تواند آن را اندازه بگیرد و با اعداد و ارقام ویژگی‌های آن را نشان دهد.
متغیر کمّی: متغیری است که اندازه‌های آن به صورت عددی بیان می‌شود: نظیر سن و بعد خانوار. این نوع متغیرها خود به دو دسته متغیرهای کمّی پیوسته و متغیرهای کمّی گسسته تقسیم می‌شوند.

متغیر کمّی پیوسته: متغیری است که هر مقدار ممکن بین حدود معینی را به خود اختصاص می‌دهد، نظیر سن، وزن، درآمد. یعنی بین هر دو مقدار آن بتوان مقادیر جدیدی را در نظر گرفت. مثلا بین ۲ کیلو و ۳ کیلو می‌توان مقادیر زیاد دیگری مثل ۲/۵ و۲/۳۷ و... را در نظر گرفت پس وزن یک متغیر کمّی پیوسته است. متغیر پیوسته اعشار می‌پذیرد. مثال دیگر، قد به عنوان یک متغیر کمّی پیوسته است چون مثلاً بین ۱ متر و ۲ متر می‌توان اندازه‌های بی‌شمار دیگری مثل ۱/۵۰ و ۱/۶۲، ۱/۸۳ و غیره را بیان کرد.

متغیر کمّی گسسته: متغیری است که فقط مقادیر معینی را می‌تواند به خود اختصاص دهد، یعنی نمی‌توان بین مقادیر آن عدد دیگری را آورد. نظیر تعداد رای دهندگان، بعد خانوار، دفعات ارتکاب جرم. مثلاً بُعد خانوار به عنوان یک متغیر گسسته مطرح است. بعد یک خانوار نمی‌تواند به صورت ۴/۲۵ یا ۴/۵ مطرح شود یا مثلاً نمی‌توانیم بگوییم تعداد فرزندان یک خانواده ۲/۳ یا ۵/۴ و... است.

پس بعد خانوار یک متغیر کمّی گسسته است. متغیر گسسته اعشار نمی‌پذیرد.

انواع متغیر از حیث تأثیر و تأثر

متغیرها از حیث این که تاثیرگذار باشند یا تاثیر پذیر، در سه وضعیت قرار می‌گیرند؛ متغیر مستقل، متغیر وابسته و متغیر همبسته.

متغیر مستقل: متغیری است که تاثیرگذار و پیش‌بینی‌کننده است.

متغیر وابسته: متغیری است که تأثیرپذیر و پیش‌بینی‌شونده است. در رابطه بین دو متغیر، متغیری را که تأثیر گذار است، مستقل و متغیری را که تاثیرپذیر است وابسته می‌نامند.

متغیرهای همبسته: دو متغیر وقتی همبسته هستند که رابطه تأثیر و تأثر با یکدیگر داشته باشند و نتوان بین آنها علت و معلول را تشخیص داد.

متغیر میانجی: اگر متغیر X از طریق Z بر Y تأثیر گذارد، آنگاه متغیر Z را متغیر میانجی یا مداخله‌گر می‌نامیم. متغیر میانجی بیانگر مکانیسم تاثیر X بر Y است. این متغیر را گاه متغیر واسط می‌نامند.

متغیر کاذب: متغیر A برای Y کاذب است وقتی که هردوی آنها تحت تاثیر متغیر X باشند. به عنوان مثال اندازه کف دست برای استعداد کلامی کودکان یك متغیر کاذب است. چون هر دو تحت تأثیر متغیر سن قرار دارند. باید توجه داشت که بین متغیر کاذب و متغیر وابسته نمی‌توان ربط علی معقولی پیدا نمود.

متغیر منکوب یا سرکوبگر (متغیر بازدارنده - پنهان کننده): هرگاه عدم کنترل یك متغیر (مثلاً (Z) موجب شود رابطه‌ی بین دو متغیر (مثلاً X و Y) ضعیف شود، آن متغیر (Z) را سرکوبگر می‌نامیم.

متغیر کنترل: هر متغیری که تأثیرش بر روابط بین متغیرهای پژوهش از طریق روش‌های آماری یا نمونه‌گیری کنترل شده باشد، متغیر کنترل نامیده می‌شود. به عنوان مثال اگر عملکرد تحصیلی دانش‌آموزان دختر و پسر را یك‌بار در مدارس دولتی و یك‌بار در مدارس غیرانتفاعی با یکدیگر مقایسه کنیم، تاثیر نوع مدرسه را بر عملکرد تحصیلی، کنترل نموده‌ایم.

امروزه با روش‌های پیشرفته آماری نظیر ضرایب همبستگی تفکیکی و نیمه تفکیکی و نیز تحلیل کوواریانس می‌توان تاثیر متغیرهای ناخواسته را کنترل نمود.

لازم به ذکر است که آنچه که به محقق کمك می‌کند تا نوع یك متغیر (مثلاً مستقل یا وابسته بودن یا کنترل یا کاذب بودن و...) را تشخیص دهد آمار نیست، بلکه نظریه‌ها و روش تحقیق است.

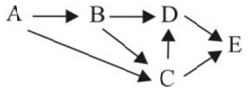
تحلیل مسیر

برای آزمون‌های مدل‌های علّی به‌کار می‌رود و مستلزم تنظیم مدل است که به صورت نمودار نشان داده می‌شود یعنی از طریق نمودار مشخص می‌کنیم کدام نوع متغیر بر متغیر دیگر تأثیر دارد.

در تحلیل مسیر با دو نوع متغیر سروکار داریم یعنی با توجه به مدل، دو نوع متغیر داریم: الف) متغیرهای درونی، ب) متغیرهای بیرونی.

متغیرهای درونی: متغیرهایی هستند که تغییرات آن‌ها از طریق متغیرهای درونی و بیرونی تبیین می‌شود و به‌عنوان متغیرهای رابطی هستند بین متغیر مستقل و وابسته.

متغیرهای بیرونی: متغیرهایی هستند که از طریق متغیرهای درونی بر متغیر وابسته تأثیر می‌گذارند. مثلاً در مدل روبه‌رو متغیرهای B، D و C متغیرهای درونی هستند و متغیر A متغیر بیرونی می‌باشد.



متغیر E که همان متغیر وابسته می‌باشد از نظر زمانی آخرین متغیر می‌باشد.

مقیاس اندازه‌گیری

اندازه‌گیری: اندازه‌گیری عبارت است از نسبت عددی دادن به یک صفت یا رویداد براساس یک قانون معین.
متغیرها از حیث مقیاس اندازه‌گیری در یك سطح نیستند و برای سنجش هر متغیری لازم است که مقیاس متناسب با آن را انتخاب نمود.
مقیاس‌ها از مهمترین ابزار پژوهش به شمار می‌آیند پس در هر سنجش ناچار باید:
۱- ویژگی مورد آزمون را به کمیّت تبدیل کرد؛
۲- درجات آن را با اعداد یا علائم خاصی مشخص ساخت؛
۳- در رده‌بندی خاصی با نام مقیاس اندازه‌گیری جای داد.
برای سنجش و اندازه‌گیری، چهار مقیاس را می‌توان نام برد:

۱- مقیاس اسمی یا طبقه‌ای

ابتدایی‌ترین مقیاس، اسمی است. در این مقیاس به تعیین طبقه‌هایی پرداخته می‌شود که افراد، اشیاء یا رویدادها را می‌توان در آنها جایگزین کرد. البته باید این طبقه‌ها ناسازگار باشند.
متغیر اسمی، متغیری است که فقط از جنبه کیفی مورد نظر است. این مقیاس برای نشان دادن وجود یا عدم وجود یك صفت در نمونه می‌باشد.
در این مقیاس اعدادی که به مقولات یک متغیر اختصاص داده می‌شود ارزش حقیقی ندارند، یعنی بیانگر بزرگی یا کوچکی مقولات نیستند و می‌توان آنها را جابه‌جا نمود .
مقیاس اسمی مطلبی درباره کم بودن یا زیاد بودن بیان نمی‌کند و دارای این ویژگی مهم است که فقط دارای طبقه‌های مشخص و متمایزی است که جنبه کیفی دارند و تنها رابطه موجود بین آنها تفاوت آنها با یکدیگر است. بسیاری از این مقیاس‌ها، دو ارزشی هستند. برای مثال جنسیّت که دارای دو ارزش مذکر و مونث است دارای مقیاس اسمی است. در مقیاس اسمی نمی‌توانیم بگوییم یک طبقه بزرگ‌تر از طبقه دیگری است و یا کوچک‌تر از دیگری است بلکه فقط می‌توانیم بگوییم که دو طبقه از یکدیگر متفاوت هستند.
متغیر اسمی یک متغیر ذاتی است یعنی خصیصه‌ای که روز به روز عوض نمی‌شود. اعدادی که در این سطح به‌کار برده می‌شوند، معنای عددی و کمّی ندارند. مناسب‌ترین آماره در این سطح، کاربرد موارد نما در این مقیاس است. نمونه انواع متغیر اسمی عبارتند از: جنسیت، ملیت، شغل، شماره پیراهن بازیکنان، شماره پلاک خانه، شماره تلفن، شماره کلاس (در این چهار مورد اخیر عدد و شماره معنای عددی ندارد) و. . .

۲- مقیاس رتبه‌ای یا ترتیبی

متغیرهای رتبه‌ای متغیرهایی‌اند که ذاتاً رتبه‌بندی شده‌اند. در این سطح محقق فقط به دسته بندی داده‌ها نمی‌پردازد بلکه آزاد خواهد بود تا آنها را با توجه به درجه‌ی اهمیت مرتب نماید.
در مقیاس ترتیبی نیز همانند مقیاس اسمی، هدف، طبقه‌بندی و تشخیص طبقات از یکدیگر است. با این تفاوت که ترتیب طبقه‌ها هم مورد توجه قرار می‌گیرد.
در این مقیاس از اعداد به منظور درجه‌بندی و رتبه‌بندی استفاده می‌شود. در این سطح اعدادی که به مقولات تعلق می‌گیرند باید از نظم و ترتیب خاصی برخوردار باشند، به عنوان مثال عددی که به مقوله متوسط تعلق می‌گیرد حتماً باید بین اعدادی باشد که به مقوله‌های کم و زیاد تعلق گرفته است.
مقیاس رتبه‌ای به منظور مرتب کردن افراد یا اشیاء از بیشترین میزان مورد اندازه‌گیری تا کمترین میزان آن به‌کار برده می‌شود. در اندازه‌گیری ترتیبی می‌توانیم تعیین کنیم که یك مشاهده (Case) بزرگ‌تر از، مساوی با، یا کوچك‌تر از بقیه مشاهدات است ولی نمی‌توانیم فاصله مشاهدات را مشخص نماییم و نیز ابتدا و انتهای آنها مشخص نیست. نظیر رضایت شغلی (خیلی راضی، راضی، متوسط، ناراضی، خیلی ناراضی).
در این مقیاس به داده‌ها بر اساس کوچک بودن یا بزرگ بودن آنها رتبه داده شده و داده‌ها با یکدیگر مقایسه می‌گردد. مثلاً مشخص می‌کنیم که یکی از دیگری بزرگ‌تر و کوچک‌تر و. . . است. به عنوان مثال تحصیلات، در مقیاس رتبه‌ای است.
در این مقیاس فاصله بین دو طبقه را نمی‌توان یکسان فرض کرد. در واقع اعداد اختصاص یافته به مقولات فاقد ارزش عددی هستند.
مثال‌های زیر در زمره متغیر ترتیبی به حساب می‌آیند: (۱۵ تا ۲۰، ۱۰ تا ۱۵، ۵ تا ۱۰، زیر ۵ سال)، (بالای۲۰ سال، ۱۵ تا ۲۰، ۱۰ تا ۱۵، ۵ تا ۱۰، زیر ۵ سال) ، (بالای۲۰ سال، ۱۵ تا ۲۰، ۱۰ تا ۱۵، ۵ تا ۱۰).
استفاده از چهار عمل ریاضی $(\times \div + -)$ در مقیاس‌های اسمی و ترتیبی امکان‌پذیر نیست.

۳- مقیاس فاصله‌ای

یکی از مشکلات مقیاس ترتیبی این است که فاصله بین مقادیر در آن نسبی است. این مشکل در متغیر فاصله‌ای وجود ندارد، زیرا در آن فاصله بین طبقات یا مقادیر یکسان است. مقیاس فاصله‌ای، مقیاسی مدرج و با درجات مساوی است. هنگامی که نتایج یک تحقیق را بر روی چنین مقیاسی وارد می‌کنیم، هم امکان رده‌بندی آنها در دو جهت (بالا و پایین) وجود دارد و هم مقایسه داده‌ها امکانپذیر است (چون درجات مساوی هستند) .

این مقیاس دارای صفر اختیاری یا قراردادی است. قراردادی بودن بدین معنی که ما خود قرارداد کرده‌ایم و برای آن یک صفر تعیین کرده‌ایم و اینکه صفر آن طبیعی نیست. مانند نقطه صفر دماسنج که ما خود نقطه‌ای را برای صفر قرارداد کرده‌ایم و صفر را برای آن تعیین نموده‌ایم. به عبارت دیگر، در مقیاس فاصله‌ای نقطه «صفر واقعی» وجود ندارد. یعنی اگر کسی در آزمون هوش نمره صفر بگیرد به این معنی نیست که او اصلا هوش ندارد.

در این مقیاس می‌توانیم اعداد را از هم کم کنیم و تفاوت بین آنها را به دست آوریم. داده‌ها همه دارای فاصله‌های مساوی هستند. با وجودی که صفر واقعی وجود ندارد، گاهی به علت سهولت و سادگی صفر قراردادی منظور می‌کنند. اندازه‌گیری درجه حرارت بر حسب فارنهایت مثالی از یک مقیاس فاصله‌ای است که نقطه صفر مطلق ندارد، بلکه دارای صفر قراردادی است و به همین دلیل عملیات ضرب و تقسیم در مورد آن انجام نمی‌گیرد و تنها می‌توان عملیات جمع و تفریق را بکار برد.

سن، بهره‌هوشی، نمره کنکور، توزیع نمرات z، هوش، عزت نفس و افسردگی در مقیاس فاصله‌ای سنجیده می‌شوند.

۴- مقیاس نسبی (کسری)

در این سطح فاصله‌ها در تمام طول مقیاس پیوستگی دارد و همچنین مبدأ سنجش در آن صفر واقعی یا صفر مطلق است. به عبارتی دقیق‌تر نمره صفر در این سطح به این معناست که فرد از آن متغیر هیچی ندارد. در این مقیاس مبدأ سنجش همیشه صفر مطلق است. در واقع مقیاس نسبی از تمام ویژگی‌های انواع مختلف مقیاس‌های پیشین برخوردار است؛ دارای بیش از یك طبقه، دارای ترتیب، مبین طول دقیق فواصل و دارای نقطه صفر حقیقی است. به عنوان مثال اگر میزان درآمد صفر باشد به این معناست که فرد اصلاً درآمد ندارد.

در این مقیاس، تمام چهار عمل اصلی ریاضی را می‌توان اعمال کرد. مقیاس نسبی قابل تبدیل به دیگر مقیاس‌هاست.

وزن، طول قد، جمعیت یک روستا، و میزان درآمد نمونه‌هایی از متغیرهای واقع در مقیاس نسبی هستند.

نکته ۳: وجه مشترک سطوح اندازه‌گیری چهارگانه اسمی، رتبه‌ای، فاصله‌ای و نسبی، در طبقه‌بندی آن‌ها است.

۲) مقیاس اسمی و ترتیبی (رتبه ای) در سطح کیفی‌اند و مقیاس فاصله‌ای و نسبی در سطح کمّی‌اند.

۳) از مقیاس نسبی به طرف مقیاس اسمی مقیاس‌ها قابل تبدیل به یکدیگرند.

۴) مقیاس اسمی قابل تبدیل به هیچ مقیاسی نیست.

۵) گرایش‌های مرکزی این مقیاس‌ها به ترتیب عبارتند از:

الف: مقیاس اسمی: مد یا نما (Mode)

ب: مقیاس ترتیبی: میانه (Median)

ج: مقیاس فاصله‌ای: میانگین (Mean)

د: مقیاس نسبی: میانگین (Mean)

کارکردهای آمار توصیفی

آمار توصیفی کارکردهای مختلفی دارد از جمله:

۱- سازمان‌دهی داده‌ها

۲- نمایش داده‌ها

۳- نشان دادن مرکزیت داده‌ها (مانند میانگین، میانه، نما و. . .)

۴- نشان دادن پراکندگی داده‌ها (مانند واریانس، انحراف معیار و. . .)

۵- نشان دادن شاخص‌های تقارن و نرمال بودن داده‌ها

۶- توصیف روابط بین داده‌ها.

سؤالات چهارگزینه‌ای کنکور سراسری فصل اول

۱- توصیف خلاصه‌ای از یک متغیر در یک نمونه چه نامیده می‌شود؟ *(سال ۸۵)*

۱) آماره ۲) پارامتر ۳) فاصله اطمینان ۴) متغیر

۲- سطوح سنجش متغیرهای قومیت، تعداد برادر و مدرک تحصیلی به ترتیب کدامند؟ *(سال ۸۸)*

۱) رتبه‌ای ـ نسبی ـ اسمی ۲) اسمی ـ نسبی ـ رتبه‌ای

۳) اسمی ـ فاصله‌ای ـ نسبی ۴) رتبه‌ای ـ فاصله‌ای ـ رتبه‌ای

۳- متغیر نوع شغل در چه سطحی سنجیده می‌شود؟ *(سال ۸۸)*

۱) اسمی ۲) نسبی ۳) فاصله‌ای ۴) رتبه‌ای

سؤالات چهارگزینه‌ای تألیفی فصل اول

٤- کدام مقیاس اندازه‌گیری صرفاً به تعیین ط بقاتی می‌پردازد که افراد، اشیاء یا حوادث را می‌توان در آنها قرار داد؟

۱) اسمی ۲) ترتیبی ۳) فاصله‌ای ۴) نسبی

٥- در کدام مقیاس اندازه‌گیری، تمام عملیات آماری و ریاضی را می‌توان انجام داد؟

۱) ترتیبی ۲) فاصله‌ای ۳) نسبی ۴) اسمی

٦- مقیاس اندازه‌گیری دو متغیر «وزن» و «بهره هوشی» در فرضیه «وزن با میزان بهره هوشی کودکان پایه اول دبستان رابطه دارد» به ترتیب کدام است؟

۱) فاصله‌ای – فاصله‌ای ۲) نسبی – نسبی ۳) نسبی – فاصله‌ای ۴) فاصله‌ای – نسبی

۷- موارد استفاده میانه بیشتر در کدام‌یک از مقیاس‌های اندازه‌گیری است؟

۱) اسمی ۲) رتبه‌ای ۳) فاصله‌ای ۴) نسبی

۸- کدام‌یک از متغیرهای زیر گسسته است؟

۱) طول عمر مفید یک رادیو ۲) نمرات آزمون آمار دانش‌آموزان

۳) فواصل بین واحدهای آموزشی یک شهر ۴) میزان زاد و ولد در شهر تهران

۹- اگر در یک مطالعه جنسیت یکی از متغیرها باشد، این متغیر چه نوع متغیری است؟

۱) وابسته ۲) پیوسته ۳) کمی ۴) ناپیوسته

۱۰- کدام‌یک از گزینه‌های زیر متغیر گسسته است؟

۱) تعداد صندلی‌ها ۲) قد مردم ۳) نمره پیشرفت تحصیلی ۴) وزن دانشجویان

۱۱- مناسب‌ترین مورد استفاده از نما زمانی است که مقیاس اندازه‌گیری به کار رفته مقیاس . . . . . باشد.

۱) اسمی ۲) ترتیبی ۳) فاصله‌ای ۴) نسبی

۱۲- با کدام روش آماری می‌توان افزایش تولید ناخالص ملی ایران را در سه سال گذشته نمایش داد؟

۱) استنباطی غیر مشروط ۲) استنباطی مشروط ۳) توصیفی ۴) توصیفی استنباطی

۱۳- هدف از کاربرد آمار استنباطی چیست؟

۱) تعمیم یافته ها به جامعه‌ی مرجع ۲) مشخص نمودن میزان خطای شاخص ها

۳) تفسیر آماره‌های محاسبه شده ۴) محاسبه شاخص ها یا آماره ها

فصل دوم

آماده‌سازی، سازمان‌دهی و نمایش داده‌ها (نمودارها)

آماده‌سازی داده‌ها

داده‌ها به هر روشی که گردآوری شده باشند داده‌های خام‌اند. اولین مرحله در تحلیل داده‌ها آماده کردن داده‌ها برای انواع تحلیل‌های مورد نظر است. اگر تحقیق، مطالعه میدانی است نوع اصلی تحلیل شامل تنظیم یادداشت‌ها برای بررسی مسأله اصلی تحقیق و سپس ارائه آنها به صورت مقاله است. ای بسا در این نوع تحلیل داده‌های کمیت‌پذیر وجود نداشته باشد.

در اکثر روش‌های دیگر برای تحلیل‌های کمّی و (احتمالاً) آزمون‌های آماری باید اطلاعات گردآوری شده اولیه (داده‌های خام) را به معادل عددی تبدیل کرد. این قسمت به نحوه تبدیل داده‌های خام بالقوه کمیت‌پذیر به اعدادی اختصاص دارد که نشان‌دهنده دامنه معنای ـ تغییر ـ داده‌های خام است. این کار برای تهیه جداول و نمودارها و شکل‌هایی که داده‌ها را به‌طور دقیقی توصیف می‌کنند و برای تبیین نحوه تأیید فرضیه‌ها براساس داده‌ها ضروری است. اگر در اندیشه آزمون آماری فرضیه‌های خود هستید باید از سنجش‌هایی برای تعیین میزان روابط استفاده کنید. چنانچه داده‌های تحقیق از نمونه احتمالی به‌دست آمده باشد می‌توان با آزمونهای آماری دیگری به استنباط‌هایی در مورد جمعیتی که نمونه از آن برگرفته، نایل آمد.

آماده کردن داده‌ها در چهار مرحله صورت می‌گیرد. مرحله اول که داده‌ها به صورت عدد **کدگذاری** می‌شوند، مرحله **کدگذاری** است. مرحله دوم که اطلاعات کدگذاری شده برای تبدیل از شکل مکتوب به شکل کامپیوتری آماده می‌شوند، مرحله **تبدیل** است. مرحله سوم که شامل فرایند عملی ورود داده‌های کدگذاری شده به کامپیوتر می‌شود، مرحله **انتقال به کامپیوتر** است.

آخرین محله، مرحله **تصحیح داده‌های وارد شده به کامپیوتر** است. در این مرحله باید صحت داده‌هایی را که وارد کامپیوتر شده‌اند وارسی کرد، چه بعید نیست در هر مرحله از پردازش داده‌ها اشتباهاتی پیش آید.

تشکیل نظام کُدگذاری

اصول کلّی

کُدگذاری داده‌ها متضمن کمّی کردن داده‌های خام برای تحلیل است. برای تشکیل نظام کُدگذاری باید از اصولی تبعیت کرد. اولین اصل این است که **کُدگذاری باید دقیقاً تعریف شده، خالی از ابهام باشد.** این امر مستلزم اتخاذ بهترین شیوه کُدگذاری برای پاسخ پرسش‌ها (متغیرها) است به گونه‌ای که معنای مفهومی که گویای متغیر تحقیق است یکسان باقی بماند. کُدگذاری را می‌توان براساس معیارهای نظری و تجربی انجام داد.

معیارهای نظری برای آزمون این است که مقیاس (یا رشته‌ای از مقیاس‌های) صفاتی که انواع تغییرات متغیری را تشکیل می‌دهند به صورت رشته‌ای از کدهای عددی درآمده باشد که دقیقاً مبین دامنه مقادیر متغیر مورد نظرند. چنین معیاری آزمون اعتبار متغیر است. معیارهای تجربی آزمون این است که نتایج سنجش متغیر تا چه حدگویای نتایج منطقی و قابل‌انتظار است. به عبارت دیگر این معیارها به تشخیص این که پاسخگویان تا چه حد

پرسش‌ها را فهمیده و به معنای پرسش شما پی برده‌اند کمک می‌کند. مسأله اصلی در اینجا این است که پاسخگویان به‌طور استوار و ثابتی به سوالات پاسخ داده باشند؛ یعنی بتوان پیش‌بینی کرد که چنانچه بار دیگر همان سوالات از آنها پرسیده شود همان پاسخ‌ها را می‌دهند. این آزمون همانا **آزمون پایایی** است. تفاوت معیارهای نظری و تجربی در کُدگذاری در این است که معیارهای نظری را می‌توان قبل از پردازش داده‌ها به کار برد اما معیارهای تجربی را فقط بعد از گردآوری داده‌ها و الگوی خاص شواهد موجود در پاسخ‌ها که راهنمای نحوه کُدگذاری است می‌توان تعیین کرد.

از جنبه نظری کدها را باید به گونه‌ای طرح کرد که گویای دامنه معنای در نظر گرفته شده برای متغیرها باشد. این دامنه معنا باید تمام تغییرات ممکن را دربر گیرد.

از جنبه تجربی، کُدگذاری داده‌ها باید دربرگیرنده تمام پاسخ‌های مختلفی باشد که پاسخگویان ارائه داده‌اند و نیز برای هنگامی است که برآنید انواع خاصی از پاسخ‌ها را ترکیب کنید. پس دومین اصل کلّی کُدگذاری این است که **حتی‌الامکان معنای واقعی پاسخ‌ها و کل دامنه تغییرات داده‌ها را کُدگذاری کرد.**

مثال بارز کُدگذاری تجربی، کُدگذاری سوالات باز است. گو این‌که در پرسش‌های باز می‌توان پاسخ‌های احتمالی را پیشاپیش در نظر گرفت اما کُدگذاری نهایی منوط به گردآوری داده‌هاست. این نوع کُدگذاری، کُدگذاری بعد از داده‌هاست و بر نتایج تجربی به‌دست‌آمده استوار است.

سومین قاعده کلّی این است که **نظام کُدگذاری را حتی‌الامکان در مرحله طرح ابزار گردآوری داده‌ها طرح‌ریزی کرد.** اگر ابزار گردآوری داده‌ها پرسش‌نامه باشد و بتوان آن را کُدگذاری کرد معیارهای نظری کُدگذاری را باید دقیقاً مراعات کرد.

چهارمین قاعده این است که **نظام کُدگذاری باید به تقلیل دفعات پردازش داده‌ها بکشد،** چه در هر بار پردازش داده‌ها اشتباهاتی رخ خواهد داد. به همین دلیل نظام کُدگذاری باید روشن و دقیق باشد.

پنجمین قاعده کُدگذاری این است که **برای تثبیت نظام کُدگذاری و تمام تصمیماتی که در مورد کُدگذاری اتخاذ شده است باید کدنامه‌ای تهیه کرد.** این کدنامه نه تنها برای راهنمای کُدگذاری لازم است، بلکه در جریان تحلیل هم برای یادآوری چیزی که متغیر می‌سنجد لازم است.

انواع استراتژی‌های کُدگذاری

پرسش‌ها معمولاً پنج گونه‌اند. گونه اول متضمن رشته‌ای از پاسخ‌های ترتیبی است مانند موافقت کامل، موافقت، مردد، مخالفت، مخالفت کامل (پرسش‌های نگرشی غالباً این‌گونه‌اند). گونه دوم دارای پاسخ‌های دوشقی است، مانند بله یا خیر و نیز پاسخ‌هایی که دال بر داشتن یا نداشتن صفتی است. پرسش «تنیس بازی می‌کنید؟ بله/ خیر» با پرسش « به کدام‌یک از این ورزش‌ها می‌پردازید؟ تنیس، شنا، بکس و ...» هم‌سنگ است. در پرسش اخیر هریک از ورزش‌ها را می‌توان مبنای یک متغیر دوشقی قرار داد مانند ۱= تنیس بازی می‌کنم و ۰= تنیس بازی نمی‌کنم.

پرسش‌های گونه سوم، پرسش‌های مطلق بوده دارای رشته‌ای از پاسخ‌های بدون ترتیب‌اند (مثلاً مذهب - کاتولیک، پروتستان، یهودی - که هیچ نظم و ترتیب منطقی ندارد). گونه چهارم پرسش‌هایی است که پاسخ آنها مستقیماً با اعداد داده می‌شود (درآمد، سن، تعداد فرزندان). گونه پنجم پرسش‌های باز است که پاسخ‌های آنها براساس محتوای پاسخ‌ها یک به یک کُدگذاری شده، سپس می‌توان آنها را از نو به طبقات عددی تبدیل کرد.

شکل نهایی کدها

شکل نهایی کدهای مکتوب همانی است که برای انتقال داده‌ها به کامپیوتر مورد استفاده قرار می‌گیرد. باز هم باید تأکید کرد که تحدید دفعات پردازش داده‌ها تأثیر بسزایی در تقلیل خطاها دارد. به همین دلیل محققان سعی می‌کنند حتی‌الامکان داده‌ها را مستقیماً از پرسش‌نامه به کامپیوتر منتقل کنند. از آنجا که این کار در پاره‌ای موارد امکان‌پذیر نیست باید به روش‌های دیگری تمسک جست:

کُدگذاری حاشیه‌ای در پرسش‌نامه. در این روش اطلاعات کُدگذاری در حاشیه پرسش‌نامه درج می‌شود به گونه‌ای که پردازش‌کننده داده‌ها فقط باید پاسخ را به محلی که مختص ستون کامپیوتری است منتقل کند (در پاره‌ای پرسش‌نامه‌ها محل پاسخ‌ها را در همان حاشیه قرار می‌دهند). کلاً کُدگذاری حاشیه‌ای فرایند کُدگذاری را ساده کرده، همسانی آن را بالا می‌برد.

استفاده از برگه کُدگذاری. هرگاه قرار باشد داده‌ها روی کارت ای بی ام (IBM) که دارای ۸۰ ستون برای ذخیره اطلاعات است پانچ شود بهتر است داده‌ها را از منبع اصلی (پرسش‌نامه) به برگه‌های مخصوص ۸۰ ستونه‌ای که عین کارت‌های ای بی ام است منتقل کرد. بدین ترتیب پانچ داده‌ها ولو آن‌که پانچ‌کن با تحقیق آشنا نباشد بسیار ساده خواهد شد. با این وصف فراموش نکنید که همین انتقال داده‌ها ممکن است باعث خطاهایی بشود که قبلاً وجود نداشته‌اند. به همین دلیل محققان ترجیح می‌دهند داده‌ها مستقیماً از پرسش‌نامه وارد کامپیوتر گردد.

دستکاری داده‌ها پیش از تحلیل

دستکاری داده‌های ناقص

در همه برنامه‌های نرم‌افزاری، روش‌هایی برای دستکاری و جابه‌جایی داده‌های ناقص وجود دارد. باید به کامپیوتر اعلام کرد که در چه مواردی داده‌های یک متغیر را ناقص به‌شمار آورد (یا با علامت مشخص کند) تا بتوان آنها را از تحلیل حذف کرد. البته اگر موردی (مثلاً پرسش‌نامه‌ای) معدودی متغیر ناقص داشته باشد به‌طور کامل از تحقیق حذف نمی‌شود، بل از کامپیوتر خواسته می‌شود متغیرهایی (پرسش‌هایی) که جواب درستی دارند باقی مانده، متغیرهایی که بی‌پاسخ مانده‌اند به‌عنوان متغیر ناقص مشخص شوند. در زبان برنامه نرم افزار به این تفکیک، **حذف جزئی داده‌های ناقص** اطلاق می‌شود. اما اگر موردی وجود داشته باشد که داده‌های همه متغیرهایش ناقص باشد به‌طور کامل از تحلیل حذف می‌شود که بدان **حذف کامل داده‌های ناقص** اطلاق می‌شود.

معمولاً با کد ۹ یا ۹۹ یا ۹۹۹ مقادیر ناقص را مشخص می‌کنند. بعد از تعیین مقادیر ناقص آنها را حذف می‌کنند، هر چند تعداد موردهای ناقص هر متغیر در جداول ثبت می‌شود.

کُدگذاری مجدد متغیرها

یکی از رایج‌ترین دستکاری‌ها در تعیین متغیرهای تحلیل، **کُدگذاری مجدد** آنهاست. عموماً هدف کُدگذاری مجدد تقلیل تعداد طبقات متغیر به تعداد معقول‌تری جهت تحلیل عددی است. وقتی برآنید رابطه دو یا چند متغیر را بررسی کنید احتمالاً باید تعداد طبقات متغیرها را تقلیل دهید تا معلوم شود اثر متغیرها بر یکدیگر قوی است یا ضعیف. اما با کُدگذاری مجدد، طبقات اولیه متغیر دگرگون خواهد شد. برای اجتناب از این اختلاط معمولاً بهتر است از متغیرهای اولیه متغیرهای جدیدی درست کرد تا هم طبقات اولیه متغیر حفظ شود و هم از متغیری با یک رشته طبقات جدید (مجدداً کُدگذاری شده) برخوردار شد.

برای کُدگذاری مجدد متغیر دو قاعده کلّی وجود دارد. یکی **ملاحظات نظری** است؛ یعنی متغیر مجدداً کُدگذاری شده به نحو بهتری معرف مفهوم متغیر تحقیق است (مثلاً در تحقیقی بر آن می‌شویم تمام غیرسفیدپوستان را در یک طبقه جای دهیم چه فقط در پی تمایز این دو هستیم). قاعده دیگر **ملاحظات تجربی** است؛ یعنی برحسب پاسخ‌های ارائه شده متغیر را مجدد طبقه‌بندی می‌کنیم (مثلاً بعد از بررسی پاسخ‌ها می‌بینیم که تعداد هر طبقه اقلیت کمتر از آنی است که بتوان به تنهایی آن را تحلیل کرد در نتیجه همه اقلیت‌ها را در یک طبقه غیرسفیدپوست ادغام می‌کنیم).

سازماندهی داده‌ها

با توجه به اینکه اطلاعات به‌دست آمده از یک تحقیق غالباً زیاد و به صورت نامنظم هستند، هر نوع نتیجه‌گیری و تعبیر در مورد آن‌ها مشکل است. بنابراین برای اینکه بتوانیم داده‌های به دست آمده را تحلیل کنیم باید آن‌ها را به صورت یک نظم منطقی درآوریم تا به صورت معنی‌دار قابل تعبیر باشند. در این راستا اولین گام، تشکیل جدول فراوانی می‌باشد که در آن مراحل طبقه‌بندی و محاسبه داده‌ها نشان داده می‌شود. پس توزیع فراوانی عبارت است از سازمان دادن به داده‌ها یا مشاهدات به صورت طبقات، همراه با فراوانی هر طبقه.

هدف کلی جداول توزیع فراوانی عبارت است از: درک بهتر داده‌های جمع‌آوری شده، کمّی ساختن واقعیت مورد مطالعه، ارائه تصویری نسبتاً دقیق از واقعیت مورد مطالعه، انعکاس واقعیت تجربی.

سازمان‌دهی داده‌ها معمولاً بر اساس جداول توزیع فراوانی صورت می‌گیرد. معمولاً داده‌های مربوط به یك متغیر به علت طبیعت ظاهری نامنظم خود گویای مطلبی درباره‌ی جامعه نیستند. برای آنکه بتوان به آنها نظم بهتری داد آنها را معمولاً در قالب جدول تنظیم می‌کنند. جداول ابزارهایی برای مهار پراکندگی داده‌ها هستند. یك جدول فراوانی معمولاً از ستون‌های زیر تشکیل می‌شود:

۱- **طبقات یا دسته‌ها:** اگر چه گاهی اوقات ستون نخست جدول می‌تواند از اعداد منفرد تشکیل شود اما اکثر اوقات دسته‌های اعداد تشکیل‌دهنده آن هستند. به طور قراردادی تعداد طبقات را بین ۱۰ تا ۲۰ انتخاب می‌نماییم. قاعده کلی این است که اگر تعداد نمره‌ها نسبتاً کم باشد (کمتر از ۱۰۰)، تعداد طبقات را نزدیك به ۱۰، ولی اگر توزیع شامل تعداد زیادی نمره است، برای مثال بیش از ۵۰۰، آنگاه طبقات را نزدیك به ۲۰ انتخاب می‌نماییم. در گام دوم فاصله بزرگ‌ترین و کوچك‌ترین عدد توزیع را محاسبه و سپس بر تعداد طبقات تقسیم می‌نماییم تا فاصله طبقات مشخص شود.

۲- **نماینده‌ی طبقات (حد میانی یا طبقه وسط طبقات):** اگر کران پایین و بالای هر طبقه را جمع و تقسیم بر ۲ نماییم نماینده‌ی طبقه به دست می‌آید.

۳- **فراوانی مطلق:** این ستون از شمارش تعداد مقادیر مربوط به هر طبقه حاصل می‌شود. جمع این ستون باید با تعداد کل نمونه برابر باشد.

۴- فراوانی نسبی: این ستون از تقسیم فراوانی مطلق هر طبقه بر جمع کل فراوانی‌ها به دست می‌آید، جمع این ستون باید ۱ شود.
۵- فراوانی درصدی: این ستون از حاصل ضرب فراوانی نسبی در عدد ۱۰۰ به دست می‌آید. جمع کل این ستون باید ۱۰۰ شود.
۶- فراوانی تجمعی درصدی: این ستون از جمع فراوانی‌های درصدی هر طبقه با فراوانی‌های درصدی طبقات ماقبل حاصل می‌شود.

طبقات	نماینده‌ی طبقات	فراوانی مطلق	فراوانی نسبی	فراوانی درصدی	فراوانی درصدی تجمعی
۱۰ تا ۲۰ هزار تومان	۱۵	۱۰	%۵	۵	۵
۲۰ تا ۳۰ هزار تومان	۲۵	۵۰	%۲۵	۲۵	۳۰
۳۰ تا ۴۰ هزار تومان	۳۵	۷۰	%۳۵	۳۵	۶۵
۴۰ تا ۵۰ هزار تومان	۴۵	۶۰	%۳۰	۳۰	۹۵
۵۰ تا ۶۰ هزار تومان	۵۵	۱۰	%۵	۵	۱۰۰
جمع	۱۷۵	۲۰۰	۱	۱۰۰	

خطای گروه‌بندی: اختلاف بین شاخص آماری محاسبه شده با استفاده از توزیع فراوانی طبقه‌بندی شده و توزیع فراوانی طبقه‌بندی نشده، خطای گروه‌بندی نام دارد.
زمانی که داده‌های یک جامعه آماری گسترده باشند، از جداول طبقاتی استفاده می‌کنیم که البته در این حالت طبقه‌بندی داده‌ها به لحاظ اقتصادی مقرون به صرفه است اما توزیع داده‌ها از نظر دقت، کمتر می‌شود.
داده‌های متغیر را می‌توان به دو صورت طبقه‌بندی کرد:
۱- طبقه‌بندی پیوسته: اگر زمانی داده‌ها به صورت پیوسته و نسبی (اعشارپذیر) بود، آنها را به صورت پیوسته طبقه‌بندی می‌کنیم.
۲- طبقه‌بندی گسسته: طبقه‌بندی گسسته برای داده‌های غیراعشاری استفاده می‌شود. داده‌هایی که از طریق شمارش به دست می‌آید (هر عددی شمارش پذیر باشد گسسته است مانند ماشین، آدم).

طبقه‌بندی پیوسته	طبقه‌بندی گسسته
۰-۵	۰-۵
۵-۱۰	۶-۱۱
۱۰-۱۵	۱۲-۱۷
۱۵-۲۰	۱۸-۲۳

در حالتی که داده‌ها به صورت گسسته طبقه‌بندی می‌شوند، امکان دارد داده‌هایی باشند که در هیچ یک از طبقات جای نگیرند برای مثال در طبقه‌بندی گسسته زیر:

۲-۴
۵-۷
۸-۱۰
۱۱-۱۳

اگر در داده‌های ما عدد ۴/۵ وجود داشته باشد، آن را در هیچ یک از طبقات نمی‌توان جای داد. به همین دلیل باید طبقات را به صورت پیوسته طبقه‌بندی کرد. برای اینکه طبقه‌بندی از حالت پیوسته به گسسته تبدیل شود باید حدود واقعی اعداد را به‌دست آوریم.
حدود واقعی عدد ۱۹ ← ۱۸/۵-۱۹/۵
حدود واقعی عدد ۱۹/۵ ← ۱۹/۴۵-۱۹/۵۵
حدود واقعی عدد ۱۲/۲۵ ← ۱۲/۲۴۵-۱۲/۲۵۵
در مثال بالا طبقه‌بندی پیوسته با احتساب حدود واقعی اعداد، اینگونه می‌شود:

۱/۵ -۴/۵
۴/۵ -۷/۵
۷/۵ -۱۰/۵
۱۰/۵ -۱۳/۵

مراحل طبقه‌بندی:

۱- مرتب‌کردن داده‌ها از کوچک به بزرگ

۲- به‌دست آوردن دامنه تغییرات (R):

کمترین عدد - بیشترین عدد = دامنه تغییرات (R) برای طبقه‌بندی پیوسته
۱+کمترین عدد- بیشترین عدد = دامنه تغییرات (R) برای طبقه‌بندی گسسته

تعیین تعداد طبقات؛ تعداد طبقات به نحوه‌ی پراکندگی داده‌ها و خواست محقق بستگی دارد و معمولاً بین ۵ تا ۱۵ طبقه توصیه می‌شود.

تعیین فاصله طبقات.

برای تعیین فاصله طبقات باید دامنه تغییرات را تقسیم بر تعداد طبقات کرد: $I = \frac{R\left(\text{دامنه طبقات}\right)}{K\left(\text{تعداد طبقات}\right)}$

تعیین حد میانی یا نماینده طبقه: معدل دو حد بالا و پایین را مرکز یا حد وسط آن طبقه می‌نامند و از طریق فرمول روبرو به‌دست می‌آید:

$$X = \frac{\text{حد بالا+ حد پایین}}{۲}$$

مثلاً در طبقه ۴/۵-۱/۵ حد میانی طبقه عدد ۳ است: $\frac{۱/۵ + ۴/۵}{۲} = ۳$

مثال ۱- در زیر نمرات درس جامعه‌شناسی یک کلاس آورده شده است. توزیع فراوانی این نمرات را با فاصله طبقاتی مناسب به‌دست آورید.

	۱۸	۱۵	۲۰	۱۲	۱٤	۱۸/۵	۱۹	۱۳	۱۷
	۱۲/۲۵	۱٦	۱۹/۵	۱۵/۷۵	۱۱	۹	۲۰	۱۸/۲۵	۱٦/۷۵

۱+ پایین‌ترین نمره - بالاترین نمره = R (دامنه تغییرات)

$R = ۲۰ - ۹ + ۱ = ۱۲$

٤= تعداد طبقات

فاصله طبقات = $i = \frac{۱۲}{۴} = ۳$

توزیع فراوانی تراکمی: فراوانی تراکمی هر طبقه مساوی است با مجموع فراوانی‌های طبقه‌های پایین‌تر و طبقه مورد نظر به عنوان مثال فراوانی تراکمی طبقه (۱۱ - ۹) همان عدد ۲، فراوانی تراکمی طبقه‌ی (۱۴- ۱۲) عدد ۶، فراوانی تراکمی طبقه‌ی (۱۷- ۱۵) عدد ۱۱، فراوانی تراکمی طبقه‌ی (۲۰- ۱۸) عدد ۱۸ یعنی همان N است.

جدول توزیع فراوانی نمرات درس جامعه شناسی گروهی از دانش‌آموزان

طبقه نمرات	فراوانی
۲۰- ۱۸	۷
۱۷- ۱۵	۵
۱۴- ۱۲	۴
۱۱- ۹	۲
	N =۱۸

نکته ۴: فراوانی تراکمی درصدی از این فرمول به‌دست می‌آید: برای این کار فراوانی تراکمی هر طبقه $\frac{CF}{N} \times ۱۰۰$ را بر N (تعداد کل اعداد) تقسیم و آنگاه ضربدر ۱۰۰ می‌کنیم.

■ نمایش داده‌ها (نمودارها)

یکی دیگر از راه‌های ارائه یافته‌ها ترسیم نمودار است. نمودار نوعی تصویر دیداری از یک مجموعه نمره یا جدول توزیع فراوانی به‌دست می‌دهد. هدف نمودار، ساده‌سازی است. نمودار به تفهیم مطلب کمک کرده و توجّه را معطوف به ماهیّت اساسی داده‌ها می‌کند. نمودارها انواع مختلفی دارند که هر کدام بر حسب شرایط خاصی طراحی می‌شوند. مزیت نمودارها نسبت به جداول، ضریب بالای سرعت انتقال مفاهیم و یافته‌ها به مخاطبان و نیز ارائه تصویر روشن‌تری از داده‌های جمع‌آوری شده است.

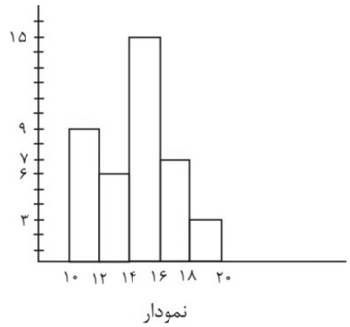
□ ۱- نمودار هیستوگرام

هیستوگرام نموداری است که در آن، فراوانی‌ها به صورت سطوحی عمودی نشان داده می‌شوند. این نمودار ویژه **متغیرهای کمّی (فاصله‌ای و نسبی) و همچنین داده‌های پیوسته** است. این نمودار از یک سری ستون‌های به هم چسبیده تشکیل می‌شود، که در آن هر ستون نشان‌دهنده طبقه‌ای از اعداد است؛ عرض هر ستون برابر فاصله طبقه و ارتفاع آن مساوی فراوانی همان طبقه است. اتصال ستون‌ها موجب می‌شود که این نمودار برای نمایش داده‌های ناشی از اجرای **متغیرهای پیوسته** وسیله مناسبی باشد. برای ترسیم این نمودار از حد بالا و پایین طبقات استفاده می‌شود.

مثال ۲- اگر نمره ریاضی۲۰۰ نفره دانشجو به شکل زیر دسته‌بندی شده باشد:

طبقات نمره	فراوانی مطلق
۱۰ تا ۱۲	۹
۱۲ تا ۱٤	٦
۱٤ تا ۱٦	۱۵
۱٦ تا ۱۸	۷
۱۸ تا ۲۰	۳

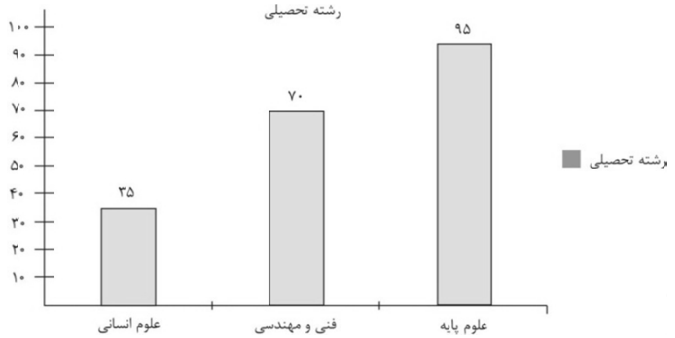
نمودار هیستوگرام آن به شکل زیر خواهد بود:



نمودار

□ ۲- نمودار ستونی یا میله‌ای

این نمودار شبیه به نمودار هیستوگرام است با این تفاوت که در این نمودار ستون‌ها مجزا از یکدیگر هستند. هر میله یا ستون نشان‌دهنده یک طبقه یا یک رده از داده‌هاست. این نمودار را می‌توان برای نمایش **داده‌های ناپیوسته** با **مقیاس‌های اسمی** که مقوله‌های جدایی از یکدیگر هستند، به آسانی مورد استفاده قرار دارد. این نمودار ویژه‌ی **متغیرهای اسمی و رتبه‌ای** است. در این نمودار باید داده‌ها **گسسته** باشند.

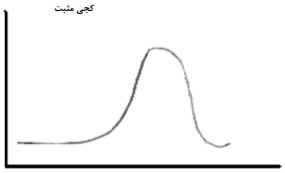


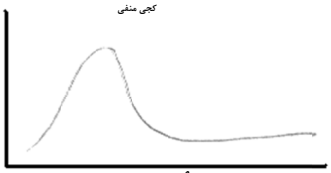
۳- نمودار چندضلعی فراوانی (نمودار چندبر یا پلی‌گون)

چنانچه در یك نمودار مستطیلی (هیستوگرام) نقطه‌ی وسط طبقات مختلف را از طریق یك خط به یکدیگر وصل کنیم، نمودار چندضلعی فراوانی پدید می‌آید. بدیهی است که اگر تعداد طبقات زیاد باشد یا به جای طبقه‌بندی، طراحی نمودار بر حسب مقادیر خام (پیوسته یا گسسته) صورت بگیرد در آن صورت به منحنی خط دست پیدا می‌کنیم. هر گاه قصد داشته باشیم دو یا چند توزیع را به صورت هندسی با همدیگر مقایسه کنیم این نمودار مناسب تر است. این نمودار برای متغیرهای کمّی (فاصله‌ای و نسبی) مورد استفاده قرار می‌گیرد. نمودار چندضلعی از تمامی نمودارهایی که به منظور توصیف توزیع‌های آماری به کار می‌رود، به دلیل سهولت در ساخت و سهولت در توصیف، بیشتر مورد استفاده قرار می‌گیرد. برای ترسیم این نمودار، از نقاط میانی طبقات استفاده می‌شود. زیرا در نمودار چند ضلعی تمام نمره‌های واقع در هر طبقه در نقاط میانی متمرکز هستند.

نمودار متقارن: نمودار چندضلعی وقتی متقارن است که نیمی از آن عیناً همانند نیمی دیگر باشد. هنگامی که یک نیمه، عیناً همانند نیمه دیگر نباشد، نمودار چندضلعی نامتقارن نامیده می‌شود؛ که این نمودار خود به دو دسته تقسیم می‌شوند؛

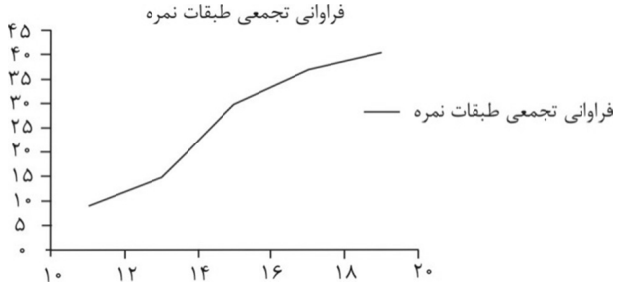
هنگامی که دنباله راست منحنی طولانی‌تر از دنباله چپ آن باشد منحنی را دارای کجی به راست یا کجی مثبت یا چولگی مثبت می‌نامند و هنگامی که دنباله چپ منحنی طولانی‌تر از دنباله راست آن باشد، توزیع دارای کجی منفی یا چولگی منفی است.





۴- نمودار چندضلعی تراکمی یا اُجایو (Ogive)

اگر مبنای منحنی خط بر اساس فراوانی تجمعی باشد نمودار حاصله اجایو خوانده می‌شود. این نمودار را از طریق قرار دادن نمره‌ها یا ویژگی مورد اندازه‌گیری بر روی محور افقی و فراوانی‌های تراکمی بر محور عمودی رسم می‌کنند. این نمودار زمانی مورد استفاده قرار می‌گیرد که پژوهشگر علاقه‌مند باشد که وضعیت نمره‌ی یک فرد را نسبت به کل توزیع فراوانی داده‌ها مشخص کند. به عبارتی تعیین کند که چند درصد از نمره‌ها بیشتر از آن طبقه و چند درصد از نمره‌ها کمتر از آن طبقه باشد. برای نشان دادن رشد و نیز نشان دادن روابط اجزاء با کل، از نمودار تجمعی استفاده می‌شود.



فراوانی تجمعی (تراکمی) را می‌توان برای به دست آوردن میانه و دامنه میان چارکی به کار برد.

۵- نمودار دایره‌ای

این نمودار برای **متغیرهایی که در سطح اسمی و رتبه‌ای قرار دارند**، مناسب است و برای سطوح فاصله‌ای و نسبی کاربرد ندارد. نمودار دایره‌ای بیشتر برای **داده‌های جمعیتی** به کار می‌رود. این نوع نمودار معمولاً برای نشان دادن رابطه اجزا با کل داده‌ها به کار می‌رود.

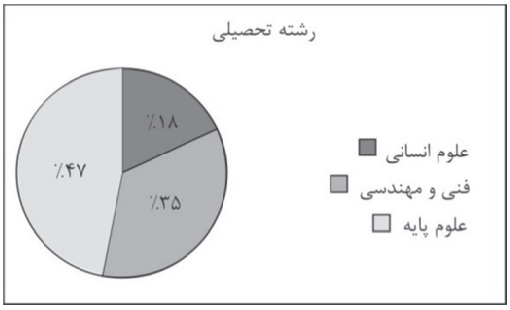
در ابتدا دایره‌ای رسم می‌شود، به طوری که محیط و سطح آن نشانگر کل (یعنی ۱۰۰%) است. سپس محیط دایره (یعنی ۳۶۰ درجه) را به n تقسیم می‌کنیم و بعد بر حسب فراوانی مطلق یا فراوانی درصدی هر طبقه یا هر بخش از داده‌ها، دایره‌ها را تقسیم‌بندی می‌کنیم. مثلاً اگر بخواهیم در یک کلاس که ۷۵درصد نمره قبولی گرفته‌اند، ۱۵ درصد نمره تجدید گرفته‌اند و ۵ درصد مردود شده‌اند را به نمایش بگذاریم، باید از نمودار دایره‌ای استفاده کنیم. برای محاسبه سهم هر یک از طبقات متغیر اسمی در نمودار دایره‌ای، فراوانی نسبی همان طبقه را در مقدار ۳۶۰ ضرب می‌کنیم:

$$S_i = \frac{F_i}{N} \times ۱۰۰^{o}$$

ترسیم نمودار دایره‌ای بر حسب هر یك از قطاع‌ها و نیز سهم هر قطاع در کل دایره را می‌توان در قالب درصد نشان داد. به عنوان مثال اگر توزیع فراوانی رشته‌های تحصیلی مختلف در یك نمونه ۲۰۰ نفری از دانشجویان به شکل زیر باشد:

علوم انسانی	فنی و مهندسی	علوم پایه
۳۵	۷۰	۹۵

نمودار دایره‌ای به شکل زیر خواهد بود:



۶- نمودار جعبه‌ای

این نمودار نشان‌دهنده چارك‌ها و حداقل و حداکثر مشاهدات است. بدین ترتیب که جعبه مشاهدات اختلاف چارك اول و سوم است. در این نمودار ابتدای جعبه چارك اول و انتهای آن چارك سوم است. خطی که جعبه را به دو قسمت تقسیم می‌کند میانه‌ی مشاهدات است.

دیاگرام‌ها: بیشتر در پدیده‌های جمعیتی بکار می‌روند و همواره دوبعدی است. فنون غیرریاضی شامل نقشه‌های جغرافیایی، نقشه‌های نقطه‌ای، نقشه‌های سایه روشن، نقشه‌های هم ارزش و نمودارهای تخیلی می‌شود.

پراکندگی به میزان تجمّع نمره‌ها حول ارزش مرکزی اشاره می‌کند. کجی به متقارن بودن یا عدم تقارن توزیع فراوانی اشاره دارد. هرگاه فراوانی‌های بزرگ‌تر در انتهای پایین متغیر و فراوانی‌های کوچک‌تر در انتهای بالای متغیر قرار بگیرند، گفته می‌شود که توزیع دارای کجی مثبت است و اگر فراوانی‌های بزرگتر در انتهای بالای متغیر و فراوانی‌های کوچک‌تر در انتهای پایین متغیر قرار گرفته باشند، گفته می‌شود که توزیع دارای کجی منفی است.

کشیدگی به پهن بودن یا بلند بودن توزیع در ارتباط با توزیع دیگر اشاره می‌کند. هرگاه توزیعی بلندتر از توزیع دیگر باشد می‌توان آن را «کشیده‌تر» و هرگاه کوتاه‌تر باشد می‌توان آن را «پهن‌تر» نامید.

توزیع طبیعی یا بهنجار کشیدگی متوسطی دارد. یعنی، از لحاظ کشیدگی بین توزیع‌های کشیده و پهن قرار می‌گیرد. گاهی نیز در برخی از توزیع‌ها، توزیع دونمایی وجود دارد.

نکته: هرگاه آزمونی دشوار را برگزار کنیم، کجی مثبت خواهد بود چون اکثر نمره ها در سمت چپ منحنی قرار می‌گیرند و هرگاه آزمونی را ساده برگزار کنیم، کجی منفی خواهد بود زیرا اکثریت نمره ها در سمت راست منحنی قرار خواهند گرفت.

سؤالات چهارگزینه‌ای کنکور سراسری فصل دوم

۱۴- کدگذاری مجدد (recoding) برای چه منظوری به کار برده نمی‌شود؟ *(سال ۹۱)*

۱) مقیاس سازی
۲) تقلیل طبقات متغیر
۳) تغییر نظم و ترتیب طبقات منظم
۴) به حداقل رساندن اثر داده‌ها و مقادیر ناقص

۱۵- کدام عبارت درباره «مجموع فراوانی‌های نسبی» یک توزیع آماری صحیح است؟ *(سال ۹۳)*

۱) $\Sigma f_i = 1$
۲) $0 \leq \Sigma f_i \leq +1$
۳) $-1 \leq \Sigma f_i \leq +1$
۴) هر مقداری را می‌تواند بپذیرد.

۱۶- برای ترسیم شکل توزیع دو متغیر که اولی «در سطح سنجش رتبه‌ای» و دومی «در سطح سنجش اسمی» اندازه‌گیری شده‌اند. به ترتیب چه نمودارهایی را پیشنهاد می‌کنید؟ *(سال ۹۴)*

۱) هیستوگرام و میله‌ای
۲) چندضلعی و دایره‌ای
۳) هیستوگرام و چندضلعی
۴) میله‌ای و دایره‌ای

سؤالات چهارگزینه‌ای تألیفی فصل دوم

۱۷- در یک دبیرستان سه رشته تجربی، ریاضی ـ فیزیک و علوم انسانی دایر است. برای نمایش درصد تعداد دانش‌آموزان کدام نمودار بهتر است؟

۱) چندضلعی
۲) خطی
۳) دایره‌ای
۴) ستونی

۱۸- برای نشان دادن روند رشد تعداد دانش‌آموزان استثنایی طی چهار سال از کدام نمودار استفاده می‌شود؟

۱) تجمعی صعودی
۲) تجمعی نزولی
۳) دایره‌ای
۴) ستونی (میله‌ای)

۱۹- نمایش هندسی نتایج آزمون‌ها را ..... می‌نامند.

۱) نمودار تراکمی
۲) نیمرخ
۳) نمودار فراوانی
۴) نمودار چندضلعی

۲۰- بهترین نمودار برای تعیین تعداد مشاهدات مساوی یا کمتر از یک مقدار معین چه نموداری است؟

۱) میله‌ای
۲) هیستوگرام
۳) ستونی
۴) فراوانی تجمعی (اُجایو)

۲۱- برای رسم نمودار چند ضلعی فراوانی تراکمی روی محور x کدام‌یک از موارد زیر درج می‌شود؟

۱) کرانه بالای هر طبقه
۲) کرانه بالا و پایین هر طبقه
۳) نماینده هر طبقه
۴) نام متغیرها یا طبقات

۲۲- اگر بخواهیم در یک کلاس تعیین کنیم چند درصد افراد از نمره ۱۲ پایین‌ترند چه نموداری مناسب است؟

۱) تراکمی درصدی
۲) تراکمی
۳) چند ضلعی
۴) درصدی

۲۳- در نمودار هیستوگرام عرض هر ستون معرف چیست؟

۱) تعداد طبقات
۲) فاصله طبقات
۳) حدود طبقات
۴) فراوانی طبقات

۲۴- در یک جدول توزیع فراوانی، برای طبقه‌ای که حد پایینی و حد بالایی آن به ترتیب ۹۹/۵ و ۱۰۹/۵ باشد نقطه‌ی میانی فاصله کدام است؟

۱) ۱۰۲/۵
۲) ۱۰۴/۵
۳) ۱۰۵/۰
۴) ۱۰۷/۰

فصل سوم

نشان دادن مرکزیت داده‌ها (پارامترهای مکان مرکزی)

اصطلاح مکان مرکزی به یک ارزش مرجع مرکزی اشاره می‌کند، که معمولاً نزدیک به بیشترین تجمع اندازه‌گیری‌ها است و تا اندازه‌ای می‌توان آن را مشخص کننده کل مجموعه نمره‌ها در نظر گرفت. شاخص‌های مرکزی به آن دسته از شاخص‌های آماری گفته می‌شود که بیانگر نقطه وسط توزیع یك متغیر هستند. در محاسبات آماری برای محاسبه ویژگی‌ها و موقعیت کلی داده‌ها، از شاخص‌های گرایش مرکزی استفاده می‌کنیم. شاخص‌های گرایش مرکزی در سه نوع نما (*mode*)، میانه (*modiau*) میانگین (*meau*) هستند که هر یک کاربرد خاص خود را دارا می‌باشند.

۱- نما یا مد (Mode)

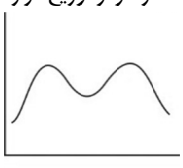
نما عبارت است از عدد، نمره یا طبقه‌ای که در توزیع فراوانی دارای **بیشترین فراوانی یا تکرار** می‌باشد. یعنی تکراری‌ترین صفت در داده‌های گروه‌بندی شده، نما است. نما ساده‌ترین شاخص گرایش مرکزی است که با ملاحظه توزیع نمره‌ها تعیین می‌شود. نما در میان شاخص‌های گرایش مرکزی ناپایدارترین (بی‌ثبات‌ترین، ضعیف‌ترین) شاخص گرایش مرکزی است.

در محاسبه نما تنها یک نمره نقش دارد و آن هم نمره‌ای است که دارای بیشترین فراوانی باشد. نما همیشه در نزدیک مرکز توزیع فراوانی قرار ندارد. به همین دلیل نمی‌توان به آن اطمینان داشت. نما، شاخصی است بی ثبات و موارد استفاده محدودی دارد. نما خصوصاً در گروه‌های کوچک دارای اعتبار نیست، زیرا میزان آن فقط تابع چند عدد است. نما را برای هر نوع توزیعی می‌توان به‌کار برد، به عبارت دیگر، نما را می‌توان در هر سطحی از اندازه‌گیری به‌کار برد، اما تنها زمانی یک شاخص مرکزی است که با **مقیاس اسمی** به‌کار برده شود.

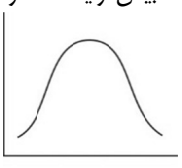
در مثال زیر نما فارس است که دارای بیشترین فراوانی است (توجه: عدد ۶۰ نما نیست بلکه فراوانی نما است).

x	F
ترک	۲۰
کرد	۳۰
فارس	۶۰

از آنجا که ممکن است نما همیشه در مرکز توزیع قرار نگیرد، بنابراین شاخص ناپایدار و بی‌ثباتی محسوب می‌گردد. مثلا در توزیع۲-۳-۳-۳-۳-۴-۴-۴-۵-۶-۷-۷-۸-۹ نما عدد۳ است اما در مرکز توزیع قرار ندارد. نما تنها شاخصی است که می‌تواند بیش از یک مقدار داشته باشد.



نمودار دو نمایی



نمودار تک نمایی

نما از بین شاخص‌های گرایش مرکزی دارای کمترین مفروضه (تعداد داده) است و همین دلیلی بر ضعیف بودن نما است.
اگر متغیر مورد بررسی از نوع فاصله‌ای یا نسبی و به صورت طبقه‌بندی‌شده باشد، نما از طریق فرمول زیر محاسبه می‌شود:

$$MO = L + \left(\frac{d۱}{d۱ + d۲}\right)h$$

که در آن:
L: حد واقعی طبقه مُد (کران پایین طبقه پیوسته نمادار)
H: فاصله طبقات (اگر داده‌ها پیوسته باشند برابرتفاضل حد بالا و پایین طبقات و اگر داده‌ها گسسته باشند از تفاضل حد واقعی کرانه بالا از کرانه پایین به‌دست می‌آید)
d_1: تفاضل فراوانی مطلق طبقه نمادار از طبقه ماقبل (طبقه کوچک‌تر)
d_2: تفاضل فراوانی طبقه مطلق یا نمادار از طبقه مابعد (طبقه بزرگ‌تر)
در داده‌های طبقه‌بندی شده نما را می‌توان از طریق فرمول زیر محاسبه کرد:

$$Mod = L + \left[\frac{f_a}{f_a - f_b}\right]$$

طبقه‌ای که بیشترین f_i را داشته باشد طبقه نمادار است.
Fa = تفاضل فراوانی مطلق طبقه نمادار با طبقه کوچک‌تر
در فرمول فوق، L حد واقعی پایینی طبقه نمایی، f_b تفاضل فراوانی مطلق طبقه نمادار با طبقه بزرگ‌تر است. i فاصله طبقاتی است.
استفاده از این فرمول زمانی مناسب است که توزیع مورد نظر بهنجار و تک‌نمایی باشد.
ابتدا باید طبقه نمادار را مشخص کنیم. یعنی طبقه‌ای که بیشترین فراوانی را دارد.

نکته ۵: اگر دو یا سه طبقه نما با فراوانی یکسان داشته باشیم باید آن طبقه‌ای را نما بگیریم که به وسط نزدیک‌تر است.

X	F
۱-۳	۲
٤-٦	٦
۷-۹	۸
۱۰-۱۲	٤

$L = ۶ / ۵$
$d۱ = ۸ - ۶ = ۲$
$d۲ = ۸ - ۴ = ۴$
$h = ۳$
$MO = ۶ / ۵ + \left(\frac{۲}{۲ + ۴}\right) \times ۲ = ۶ / ۵ + \left(\frac{۱}{۳}\right) \times ۲ = ۷ / ۵$

نما را وقتی محاسبه می‌کنیم که:
الف) برآورد سریعی از اندازه‌های مرکزی توزیع لازم باشد.
ب) برآورد تقریبی از اندازه‌های مرکزی توزیع لازم باشد.
ج) بخواهیم بدانیم کدام نمره بیشتر تکرار شده است.
د) مقیاس ما اسمی باشد.

۲- میانه

میانه نقطه وسط در توزیع نمره‌هاست و نقطه‌ای است که نیمی از توزیع نمره‌ها در بالای آن و نیمی دیگر در پایین آن قرار دارد. به میانه نقطه ۵۰% نیز می‌گویند. اولین قدم در محاسبه میانه مرتب کردن اعداد می‌باشد. سپس نمره‌ای که توزیع را به دو نیمه تقسیم می‌کند، به عنوان میانه انتخاب می‌نمائیم. اگر توزیعی را به صورت۶، ۵، ۵، ۴، ۳، ۳، ۲ داشته باشیم در این توزیع عدد ۴ میانه است.

نکته ۶: در صورتی که تعداد نمره‌ها در توزیع فرد باشد، میانه عددی است که در وسط قرار دارد، اما در صورتی که تعداد نمره‌ها زوج باشد، میانه عبارت است از معدل دو نمره‌ای که در وسط واقع می‌شوند.

چنانچه نمره یا عددی که توزیع را به دو قسمت تقسیم می‌کند تکراری باشد، میانه به صورت زیر محاسبه می‌شود. به نمره‌های زیر توجه کنید:

۴ ۵ ۶ ۷ ۷ ۷ ۸ ۹ ۹ ۱۰

چون ۱۰ عدد در توزیع وجود دارد. میانه نقطه‌ای است که ۵ عدد در بالای آن و ۵ عدد در پایین آن واقع می‌شود. به منظور تعیین نقطه‌ای که توزیع را به دو قسمت تقسیم کند، فرض بر این است که سه تا عدد ۷ در حدود واقعی عدد ۷، یعنی ۷/۵ - ۶/۵ پراکنده شده‌اند، این مفروضه با این فرض آماری که اعداد به صورت پیوسته توزیع شده‌اند هماهنگی دارد. بدین ترتیب میانه در محدوده ۷/۵- ۶/۵ قرار دارد.

سه عدد در زیر ۶/۵ (حد پایین ۷) قرار دارند، بنابراین برای یافتن میانه از بین کلیه اعدادی که در محدوده ۷/۵- ۶/۵ قرار دارند، فقط به دو عدد آنها نیاز داریم، چون سه تا هفت وجود دارد و ما به دوتای آنها نیاز داریم، بنابراین دو سوم فاصله‌ای را که بین ۷/۵- ۶/۵ قرار دارد انتخاب می‌کنیم. عدد ۲ که در صورت کسر قرار دارد عبارت است از تعداد اعدادی که لازم داریم و عدد ۳ نشان‌دهنده تعداد عدد ۷ است که در حدود واقعی ۷/۵- ۶/۵ قرار دارند.

بنابراین میانه برابر است با: $۶/۵ + \frac{۲}{۳} = ۶/۵ + ۰/۶۷ = ۷/۱۷$

در میانه:

الف: اگر توزیع فراوانی یك متغیر طبقه‌بندی نشده باشد: ۱- ابتدا مقادیر را مرتب می‌کنیم، ۲- سپس از طریق فرمول زیر محل میانه را به دست می‌آوریم:

$$md = \frac{n+۱}{۲}$$

در این اعداد ۷، ۶، ۵، ۴، ۳، ۲ محل میانه ۳/۵ است.

$$md = \frac{۶+۱}{۲} = ۳/۵$$

این فرمول محل میانه را تعیین می‌کند نه خود میانه را.

اما برای اینکه خود میانه را پیدا کنیم باید دو عددی را که محل میانه یعنی عدد ۳/۵ در بین آنهاست را پیدا کنیم. این دو عدد ۴ و ۵ هستند. و این دو عدد را جمع و تقسیم بر۲ می‌کنیم؛ عدد حاصل میانه است.

$$md = \frac{۴+۵}{۲} = ۴/۵$$

ب: اگر مقادیر یك توزیع به صورت طبقه‌بندی‌شده باشد، مقدار میانه را از فرمول زیر به دست می‌آورند (البته برای توزیع غیرطبقه‌ای نیز می‌توان از این فرمول استفاده کرد):

$$md = L + \left(\frac{\frac{n}{۲} - CF_{i-۱}}{f_i} \right) h$$

که در آن:

L: حد واقعی طبقه میانه (کرانه پایین طبقه‌ای که میانه در آن قرار دارد)

CF: فراوانی تراکمی ماقبل طبقه میانه

f: فراوانی مطلق طبقه‌ی میانه

h: فاصله طبقاتی(در حالتی که توزیع غیرطبقه‌ای است فاصله طبقاتی ۱ است)

n: تعداد نمرات یا Σf_i

اولین قدم در محاسبه میانه پیدا کردن طبقه میانه است. برای پیدا کردن طبقه میانه ابتدا می‌بایست ستون فراوانی تراکمی را تشکیل داد و بعد تعداد فراوانی‌ها را تقسیم بر ۲ کرد ($\frac{N}{۲}$) و بعد $\frac{N}{۲}$ را به ستون CF (فراوانی تراکمی) ارجاع نمود. اگر $\frac{N}{۲}$ با CF یکی بود، همان طبقه، طبقه میانه است وگرنه اولین فراوانی تراکمی بزرگ‌تر از آن طبقه میانه خواهد بود. در این مثال مقدار میانه را با استفاده از فرمول بالا به‌دست می‌آوریم.

X	f (فراوانی)	CF (فراوانی تراکمی)
۱-۳	۲	۲
۴-۶	۶	۸
۷-۹	۸	۱۶
۱۰-۱۲	۴	۲۲

$\frac{N}{۲}=\frac{۲۰}{۲}=۱۰$

$L=۶/۵$

$CF=۸$

$f=۸$

$h=۳$

$$md=L+(\frac{\frac{n}{۲}-CF_{i-۱}}{f_i})h=۶/۵+\left(\frac{\frac{۲۰}{۲}-۸}{۳}\right)۳=۷/۲۵$$

مثال ۳- برای محاسبه میانه در شرایطی که داده ها طبقه بندی نشده‌اند میانه‌ی جدول روبه‌رو را محاسبه کنید.

X	F	CF
۱۰	۱	۲۰
۹	۲	۱۹
۸	۳	۱۷
۷	۵	۱٤
٦	٤	۹
۵	۲	۵
٤	۵	۳
۳	۱	۱

برای جواب دادن ابتدا ستون فراوانی تجمعی را به جدول اضافه می‌کنیم. آنگاه N یعنی ۲۰ را تقسیم بر دو می‌کنیم می‌شود ۱۰. می‌آییم در ستون فراوانی تجمعی عددی که بزرگ‌تر یا مساوی ۱۰ است را می‌یابیم؛ عدد مذکور ۱٤ است. آنگاه برای یافتن L به x همین طبقه‌ی ۱٤ مراجعه کرده و حد پایین عدد مربوط (در اینجا ۷) را می‌یابیم. آنگاه اعداد را در فرمول جاگذاری می‌کنیم.

$$Md=\quad L+\frac{\frac{N}{۲}-CF}{F}(i)\Rightarrow \qquad ۶/۵+\frac{\frac{۲۰}{۲}-۹}{۵}=۶/۷$$

X	F	CF
۱۰	۱	۲۰
۹	۲	۱۹
۸	۳	۱۷
۷	۵	(۱٤)
٦	٤	۹
۵	۲	۵
٤	۵	۳
۳	۱	۱

→ N

برای محاسبه میانه در توزیع فراوانی گروهی نیز از فرمول زیر استفاده می‌شود:

$$Md=L+\frac{\frac{n}{۲}-cf}{f}(h)$$

ویژگی‌های میانه md:

الف) اندازه یا حجم واحدهای اندازه‌گیری در میانه تأثیری ندارد.

ب) میانه نسبت به اعداد بزرگ یا کوچک حساس نیست.

ج) مجموع قدر مطلق انحراف نمره‌ها از میانه، کوچکتر یا مساوی قدر مطلق انحراف‌های نمره‌ها از هر عدد دیگری است.

به عبارت دیگر، مجموع قدر مطلق‌های تفاوت‌های مقادیر مختلف از میانه، در کمترین حد است.

$$\sum |x_i - md| = \min \cdots or \cdots \sum f_i |x_i - md| = \min$$

د) در هر جامعه آماری فقط یك میانه وجود دارد.

ه) وقتی همه مقادیر معلوم نیستند باز هم میانه قابل محاسبه است.

و) اگر در یك توزیع مقادیر پرت و دور افتاده وجود داشته باشد، بهتر است از میانه به عنوان شاخص مرکزی استفاده شود. میانه بر دیگر شاخص‌ها از جمله میانگین برتری دارد، چرا که میانگین متأثر از داده‌های پرت می‌باشد (یعنی میانگین داده‌های پرت را به سمت خود می‌کشد).

ز) میانگین بر مجموع، مد بر تکرار و میانه بر ترتیب مبتنی است. و چون میانه بر ترتیب مبتنی است بنابراین بیشتر زمانی مورد استفاده قرار می‌گیرد که سطح سنجش داده‌ها ترتیبی باشد.

میانه را وقتی محاسبه می‌کنیم که:

الف) وقت کافی برای محاسبه میانگین نداشته باشیم.

ب) توزیع دارای چولگی قابل ملاحظه باشد، یعنی وقتی که یک یا چند نمره در حد افراط و تفریط قرار گیرد.

ج) بخواهیم بدانیم که نتایج اندازه‌گیری در نیمه بالا یا در نیمه پایین توزیع قرار می‌گیرد.

د) توزیع نمره‌ها باز یا ناقص باشد.

نکته ۷: مهم! در محاسبه میانه نباید مرتب کردن اعداد از کوچک‌تر به بزرگ‌تر را فراموش کنیم.

۳- میانگین

میانگین پایدارترین، معتبرترین، باثبات‌ترین و دقیق‌ترین شاخص گرایش مرکزی است. میانگین بهترین شاخص گرایش مرکزی است، زیرا برای محاسبه آن از تمامی تصاویر توزیع استفاده می‌شود. برای اغلب توزیع‌های فراوانی درآمد، مناسب‌ترین شاخص تمایل مرکزی میانگین می‌باشد. میانگین متأثر از داده‌های پرت است.

میانگین یا میانگین حسابی عبارت است از جمع تمام نمرات یك توزیع فراوانی تقسیم بر تعداد آنها.

الف: اگر توزیع یك متغیر به صورت مقادیر منفرد باشد، مقدار میانگین را از طریق فرمول زیر محاسبه می‌نماییم:

$$\overline{x} = \frac{\sum x_i}{n}$$

مثلا اگر داده‌ها در یک توزیع به صورت: ۲، ۳، ۴ ، ۵ ، و۱ باشند داریم:

$$\overline{x} = \frac{\sum x_i}{n} = \frac{(۲+۳+۴+۵+۱)}{۵} = ۳$$

ب: اگر توزیع یك متغیر دارای فراوانی باشد، مقدار میانگین از طریق فرمول زیر محاسبه می‌شود:

$$\overline{x} = \frac{\sum f_i x_i}{n}$$

X_i	F_i	$x_i F_i$
۲	۳	۶
۳	۳	۹
۴	۲	۸
۵	۱	۵

$$n = \Sigma f_i n$$

$$\sum f_i x_i = (۲/۳) + (۳/۳) + (۴/۲) + (۵/۱) = ۲۸$$

$$\overline{x} = \frac{\sum f_i x_i}{n} = \frac{۲۸}{۹} = ۳/۱۱$$

X	Fi	نماینده طبقات	نماینده طبقات $Fi \times Xi$
۰-۲	۲	۱	۲
۲-۴	۴	۳	۱۲
۴-۶	۳	۵	۱۵
۶-۸	۱	۷	۷

ج: اگر توزیع یك متغیر به صورت طبقه‌بندی‌شده باشد، ابتدا نماینده‌ی طبقات را محاسبه نموده و سپس مقدار میانگین از طریق فرمول فوق محاسبه می‌شود.

$\overline{x} = \frac{\sum f_i x_i}{n} = \frac{۳۶}{۱۰} = ۳/۶$ نماینده طبقات $= \frac{\text{کرانه پایین طبقه + کرانه بالای طبقه}}{۲}$

میانگین میانگین‌ها

اگر m_1 میانگین F_1، m_2 میانگین F_2، m_3 میانگین F_3 و... و m_k میانگین F_k باشد، میانگین جمع این اعداد برابر است با:

$$\overline{x} = \frac{F_۱m_۱ + F_۲m_۲ + F_۳m_۳ + ... + F_k m_k}{F_۱ + F_۲ + F_۳ + ... + F_k}$$

فرمول میانگین میانگین‌ها عبارت است از: $\overline{x} = \frac{\Sigma \overline{x}_i}{\Sigma n_i}$ و در صورتی که هر یک از میانگین‌ها فراوانی‌دار باشند: $\overline{x} = \frac{\Sigma \overline{x}_i f_i}{\Sigma n_i}$

	گروه ۱	گروه ۲	گروه ۳
(میانگین)	۱۰	۳۰	۲۰
f	۵	۴	۳
$f_i \ \overline{x}$	۵۰	۱۲۰	۶۰

$$\overline{x} = \frac{\Sigma \overline{x}_i f_i}{n} = \frac{(۵۰ + ۱۲۰ + ۶۰)}{۱۲} = \frac{۲۳۰}{۱۲} = ۱۹/۱۶$$

ویژگی‌های میانگین:

۱- در هر جامعه‌ی آماری فقط یك میانگین وجود دارد و میانگین مرکز ثقل داده‌هاست.

۲- مقادیر بزرگ و کوچك به سهم خود در میانگین سهیم‌اند.

۳- میانگین تنها شاخصی است که اگر به جای کلیه داده‌ها قرار گیرد مجموع آن تغییر نخواهد کرد. یعنی: $\sum x_i = N.\overline{x}$

۴- مجموع انحراف نمره‌ها یا مقادیر از میانگین صفر است.

$$\sum (x_i - \overline{x}) = ۰ \text{ ----or---- } \sum f_i (x_i - \overline{x}) = ۰$$

۵- مجموع مجذورات انحرافات داده‌ها از میانگین، کوچک‌تر یا مساوی با مجموع مجذور انحراف داده‌ها از هر عدد دیگری به غیر از میانگین است. بنابراین مجموع مجذورات انحراف مقادیر از میانگین برابر با حداقل مقدار ممکن است. به این خاصیت میانگین، خاصیت **حداقل مربعات** می‌گویند.

$$\sum (x_i - \overline{x})^۲ = \min \text{---or----} \sum f_i (x_i - \overline{x})^۲ = \min$$

۶- میانگین عدد ثابت خود عدد ثابت است. یعنی: $\overline{a} = a$

۷- اگر مقادیر در یك عدد معین ضرب شود میانگین در همان عدد ضرب می‌شود. $\overline{x^* a} = \overline{x^*}.a$

۸- اگر مقادیر بر یك عدد معین تقسیم شود، میانگین بر همان عدد تقسیم می‌شود. $\overline{\left(\frac{x^*}{a}\right)} = \frac{\overline{x}}{a}$

۹- اگر مقادیر با یك عدد معین جمع یا تفریق شوند میانگین به همان مقدار اضافه یا کم می‌شود.

$$\overline{x + a} = \overline{x} + a \text{ ------or------ } \overline{x - a} = \overline{x} - a$$