



آمار و ریاضی

پژوهش علوم اجتماعی

مؤلفان: پریسا حاج کریمی، مسین صنعتی

ویراستار علمی: سینا احمدی

سری کتاب‌های کمک آموزشی کارشناسی ارشد



حاج کریمی، پریسا

آمار و ریاضی رشته پژوهش علوم اجتماعی / پریسا حاج کریمی، حسین صنعتی / ویراستار علمی: سینا احمدی
مهر سبحان، ۹۹

۳۴۸ ص: جدول، نمودار (آمادگی آزمون کارشناسی ارشد پژوهش علوم اجتماعی)

ISBN: 978-600-334-073-2

فهرست نویسی بر اساس اطلاعات فیبا.

فارسی - چاپ سوم

۱- آمار و ریاضی

۲- آزمونها و تمرینها (عالی)

۳- آزمون دوره‌های تحصیلات تکمیلی

۴- دانشگاهها و مدارس عالی - ایران - آزمونها

پریسا حاج کریمی، حسین صنعتی

ج - عنوان

۲۳ ت ۳۴۷ الف/۳۵۳ LB

رده‌بندی دیویی: ۳۷۸/۱۶۶۴

شماره کتابشناسی ملی: ۱۱۲۶۷۶

آمار و ریاضی

نام کتاب:

پریسا حاج کریمی، حسین صنعتی

مؤلفان:

مهر سبحان

ناشر:

سوم / ۹۹

نوبت و تاریخ چاپ:

جلد ۲۰۰

تیراژ:

۸۲۰/۰۰۰

قیمت:

ISBN: 978-600-334-073-2

شابک:

انتشارات مهر سبحان: خیابان ولیمصر، بالاتر از تقاطع مطهری، روبروی قنادی هتل بزرگ تهران،
جنب بانک ملی، پلاک ۲۰۵۰
تلفن: ۴-۸۸۱۰۰۱۱۳

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

مقدمه ناشر

آیا آنانکه می‌دانند با آنانکه نمی‌دانند برابرند؟ (قرآن کریم)

پس از حمد و سپاس و ستایش به درگاه بی‌همتای احدیت و درود بر محمد مصطفی، عالی نمونه بشریت که در تاریخ دور تاریخ، بنا به فرمان نافذ صمدیت از میان مردمی برخاست که خود بودند در پست‌ترین حد توحش و ضلال و بربریت و آنگاه با قوانین شامل خویش هم ایشان را راهبری نمود و رهانید از بدویت و استعانت جوییم از قرآن کریم، کتابی که هست جاودانه و بی‌نقص تا ابدیت.

کتابی که در دست دارید آخرین ویرایش از مجموعه کتب خودآموز مؤسسه آموزش عالی آزاد ماهان است که بر مبنای خلاصه درس و تأکید بر نکات مهم و کلیدی و تنوع پرسش‌های چهار گزینه‌ای جمع‌آوری شده است. در این ویرایش ضمن توجه کامل به آخرین تغییرات در سرفصل‌های تعیین شده جهت آزمون‌های ارشد تلاش گردیده است که مطالب از منابع مختلف معتبر و مورد تأکید طراحان ارشد با ذکر مثال‌های متعدد بصورت پرسش‌های چهار گزینه‌ای با کلید و در صورت لزوم تشریح کامل ارائه گردد تا دانشجویان گرامی را از مراجعه به سایر منابع مشابه بی‌نیاز نماید.

لازم به ذکر است شرکت در آزمون‌های آزمایشی ماهان که در جامعه آماری گسترده و در سطح کشور برگزار می‌گردد می‌تواند محک جدی برای عزیزان دانشجو باشد تا نقاط ضعف احتمالی خود را بیابند و با مرور مجدد مطالب این کتاب، آنها را برطرف سازند که تجربه سال‌های مختلف موکد این مسیر به عنوان مطمئن‌ترین راه برای موفقیت می‌باشد.

لازم به ذکر است از پورتال ماهان به آدرس www.mahanportal.ir می‌توانید خدمات پشتیبانی را دریافت دارید.

و نیز بر خود می‌بالیم که همه ساله میزان تطبیق مطالب این کتاب با سؤالات آزمون‌های ارشد- که از شاخصه‌های مهم ارزیابی کیفی این کتاب‌ها می‌باشد- ما را در محضر شما سربلند می‌نماید.

در خاتمه بر خود واجب می‌دانیم که از همه اساتید بزرگوار و دانشجویان ارجمند از سراسر کشور و حتی خارج از کشور و همه همکاران گرامی که با ارائه نقطه نظرات سازنده خود ما را در پربارتر کردن ویرایش جدید این کتاب یاری نمودند سپاسگزاری نموده و به پاس تلاش‌های بی‌چشمداشت، این کتاب را به محضرشان تقدیم نماییم.

مؤسسه آموزش عالی آزاد ماهان

معاونت آموزش

مقدمه مؤلف

دوست گرامی!

با عرض سلام و احترام

درس آمار و ریاضی به عنوان یکی از مهم‌ترین دروس رشته پژوهش علوم اجتماعی و همچنین بخشی از مباحث درس روش تحقیق در رشته‌های پژوهش، مجموعه علوم اجتماعی و مطالعات زنان محسوب می‌شود. سعی ما در این مجموعه بر این بوده است که مهمترین مطالب را از منابع آزمونی این درس فراهم آوریم و با زبانی روان تشریح کرده و تقدیم شما نماییم. هریک از بخش‌ها و یا فصل‌های این کتاب دارای سوالات چهارگزینه‌ای است که هدف از طرح آنها کمک به خواننده در افزایش تسلط بر مطالب مربوطه از زوایای مختلف است. در پایان هر کدام از بخش‌ها نیز سوالات کنکوری مربوط به همان بخش درج شده و به صورت مشروح پاسخ داده شده‌اند.

امتیاز این مجموعه در این است که مطالب مفید غالب کتاب‌های مشابه موجود در بازار را در خود جمع آورده و در عین حال سعی در پرهیز از اشکالات و نقایص آنها داشته است. به علاوه، در تدوین این کتاب آخرین تغییرات در منابع طرح سوالات آزمونی لحاظ شده و مطالب آزمونی مفید موجود در منابع جدید نیز درج شده‌اند. اما در همین خصوص، یک نکته لازم به ذکر است و آن اینکه در آزمون سه سال اخیر رشته پژوهش علوم اجتماعی، برخلاف سال‌های پیشین، اکثر سوالات درس آمار و ریاضی از مباحث ریاضی طرح شده است. لیکن ما از آوردن مطالب مربوط به ریاضی در این کتاب اجتناب ورزیدیم، چرا که برای گنجاندن مطالب ریاضی مربوط به سوالات کنکور در این کتاب، لازم بود حجم وسیعی از مباحثی را که مورد نیاز داوطلب نیست، به عنوان مقدمه درج نماییم. به عنوان مثال برای مطرح کردن مطالب مربوط به سوالات دیفرانسیل و انتگرال که در سالهای گذشته مورد سوال قرار گرفته‌اند، چاره‌ای جز درج مباحث مربوط به حد و پیوستگی و توابع و مشتقات و... به عنوان مقدمه نیست. اما درج این مباحث به‌طور قطع از توان این کتاب و حوصله داوطلبان رشته پژوهش علوم اجتماعی خارج است. با توجه به نکات فوق، تصمیم مولفان این مجموعه بر این شد که در فصل مربوط به مباحث ریاضی (فصل هشتم) به درج سوالات کنکوری این مبحث و ارائه پاسخ‌های تشریحی این سوالات اکتفا شود و خوانندگان گرانقدر این مجموعه، در صورت تمایل به مطالعه مباحث ریاضی، به کتاب ریاضی رشته مدیریت موسسه ماهان ارجاع داده شوند. 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.sanati@gmail.com

عنوان	صفحه
بخش اول: مفاهیم اساسی آماری	۷
فصل اول: مفاهیم اساسی آماری	۹
سوالات چهار گزینه‌ای فصل اول	۱۷
پاسخنامه سوالات چهار گزینه‌ای فصل اول	۱۸
بخش دوم: آمار توصیفی و کارکردهای آن	۱۹
فصل دوم: سازمان‌دهی و نمایش داده‌ها	۲۱
سوالات چهار گزینه‌ای فصل دوم	۳۰
پاسخنامه سوالات چهار گزینه‌ای فصل دوم	۳۱
فصل سوم: نشان دادن مرکزیت داده‌ها	۳۳
سوالات چهار گزینه‌ای فصل سوم	۵۱
پاسخنامه سوالات چهار گزینه‌ای فصل سوم	۵۵
فصل چهارم: پارامترهای پراکندگی	۵۷
سوالات چهار گزینه‌ای فصل چهارم	۷۲
پاسخنامه سوالات چهار گزینه‌ای فصل چهارم	۷۵
سوالات کنکوری بخش دوم	۷۸
پاسخنامه سوالات کنکوری بخش دوم	۹۳
بخش سوم: آمار استنباطی و کارکردهای آن	۱۱۳
فصل پنجم: آمار استنباطی (برآورد و آزمون فرضیه)	۱۱۵
سوالات چهار گزینه‌ای فصل پنجم	۱۵۷
پاسخنامه سوالات چهار گزینه‌ای فصل پنجم	۱۶۲
فصل ششم: تخمین آماری پارامترهای جامعه	۱۶۵
سوالات چهار گزینه‌ای فصل ششم	۱۸۰
پاسخنامه سوالات چهار گزینه‌ای فصل ششم	۱۸۸
سوالات کنکوری بخش سوم	۱۹۷
پاسخنامه سوالات کنکوری بخش سوم	۲۰۰
بخش چهارم: احتمال	۲۰۳
فصل هفتم: احتمال	۲۰۵
سوالات چهار گزینه‌ای فصل هفتم	۲۲۵
پاسخنامه سوالات چهار گزینه‌ای فصل هفتم	۲۳۱
سوالات کنکوری بخش چهارم	۲۳۸
پاسخنامه سوالات کنکوری بخش چهارم	۲۴۰

۲۴۳	بخش پنجم: ریاضی
۲۴۵	فصل هشتم: ریاضی
۲۴۷	سوالات کنکور ریاضی بخش پنجم
۲۵۰	پاسخنامه سوالات کنکور ریاضی بخش پنجم
۲۵۵	فصل نهم: نکات آماری مهم
۲۷۰	سوالات چهار گزینه‌ای فصل نهم
۲۷۶	پاسخنامه سوالات چهار گزینه‌ای فصل نهم
۲۸۱	سوالات چهار گزینه‌ای کنکور سراسری سال ۸۷-۹۳
۳۴۳	منابع

بخش اول

مفاهیم اساسی آماری

عناوین اصلی

❖ فصل اول: مفاهیم اساسی آماری

مفاهیم اساسی آماری

عناوین اصلی

- ❖ تعریف علم آمار
- ❖ آمار توصیفی و آمار استنباطی
- ❖ تعریف برخی از مفاهیم آماری
- ❖ متغیر و انواع آن
- ❖ مقیاس اندازه‌گیری
- ❖ کارکردهای آمار توصیفی

فصل اول

مفاهیم اساسی آماری

تعریف علم آمار

آمار علمی است که با استفاده از فنون و روش‌های علمی و ریاضی به جمع‌آوری، طبقه‌بندی و پیش‌بینی اطلاعات کمی 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 به دست 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 فعالیت 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 آمارها و تعیین ارتباط بین اطلاعات می‌پردازد، آمار توصیفی می‌نامند. آمار توصیفی برای توصیف گروهی که مورد مطالعه‌ی محقق قرار گرفته‌اند، به کار می‌رود. هدف در آمار توصیفی، تقلیل، توصیف، تلخیص و نمایش روابط بین متغیرها است. در واقع هنگامی که توده‌ای از اطلاعات کمی (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 به کار برده می‌شود. شاخصهایی که در آمار توصیفی به کار برده می‌شوند عبارتند از: مد، میانگین، معدل و همبستگی. توجه شود که روشهای آمار توصیفی همیشه برای تعیین و بیان ویژگی‌ها 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 یک 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 آمارها 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 نمونه‌ای است.

هدف عمده در آمار استنباطی، استنباط پارامتر از آماره است، و در واقع، در آمار استنباطی آماره‌ها را به پارامترها تعمیم می‌دهیم.

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

نکته: 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 ناریب: یک نمونه‌ی تصادفی یا ناریب (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 یک 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 **متغیر** به خصوصیتی اشاره دارد که به وسیله‌ی آن اعضای یک گروه یا مجموعه از هم متمایز می‌شوند. **متغیر کمی** 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 یک 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 مانع‌الجمع 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 مانع‌الجمع بودن هم یعنی اینکه مثلاً اگر بگوییم مسیحی، مسلمان، شیعه، زرتشتی، اشتباه است چون دو طبقه آن یعنی شیعه و مسلمان با هم مانع‌الجمع نمی‌باشند و شیعه‌ها، مسلمان هستند.

این متغیرها قابل تبدیل شدن 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 هم چنان 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 ۱ و ۲ جنبه‌ی ریاضی ندارند و صرفاً جهت کدگذاری بکار 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 کمی: متغیری است که اندازه‌های آن به صورت عددی بیان می‌شود: نظیر سن و بعد خانوار. این نوع متغیرها خود به دو دسته متغیرهای کمی پیوسته و متغیرهای کمی گسسته تقسیم می‌شوند.

متغیر کمی 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. مثلاً بین ۲ کیلو و ۳ کیلو می‌توان مقادیر زیاد دیگری مثل ۲/۵ و ۲/۳۷ و... را در نظر گرفت پس وزن یک متغیر کمی پیوسته است. متغیر پیوسته اعشار می‌پذیرد. مثال دیگر، قد به عنوان یک متغیر کمی پیوسته است چون مثلاً بین ۱ متر و ۲ متر می‌توان اندازه‌های بی‌شمار دیگری مثل ۱/۵۰ و ۱/۶۲ و ۱/۸۳ و غیره را بیان کرد.

متغیر کمی 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 بعد 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 کمی گسسته است. متغیر گسسته اعشار نمی‌پذیرد.

انواع متغیر از حیث تأثیر و تأثر

متغیرها از حیث این که تأثیرگذار باشند یا تأثیر پذیر، در سه وضعیت قرار می‌گیرند؛ متغیر مستقل، متغیر وابسته و متغیر همبسته. **متغیر مستقل:** متغیری است که تأثیرگذار و پیش‌بینی‌کننده است.

متغیر وابسته: متغیری است که تأثیرپذیر و پیش‌بینی‌شونده است. در رابطه بین دو متغیر، متغیری را که تأثیرگذار است، مستقل و متغیری را که تأثیرپذیر 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 تأثیر X بر Y است. این متغیر را گاه **متغیر واسط** می‌نامند.

متغیر کاذب: متغیر A برای Y کاذب است وقتی که هر دوی آنها تحت تأثیر متغیر X باشند. به عنوان مثال اندازه کف دست برای استعداد کلامی کودکان یک 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 یک 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 یکبار در مدارس دولتی و یک بار در مدارس غیر انتفاعی با یکدیگر مقایسه کنیم، تأثیر 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 تأثیر متغیرهای ناخواسته را کنترل نمود.

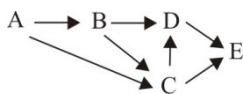
لازم به ذکر است که آنچه که به محقق کمک می‌کند تا نوع یک متغیر (مثلاً مستقل یا وابسته بودن یا کنترل یا کاذب بودن و...) را تشخیص دهد آمار نیست، بلکه نظریه‌ها و روش تحقیق است.

تحلیل مسیر

برای آزمون‌های مدل‌های علی بکار 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 به عنوان 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 روبرو 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 یک 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 یک 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 جنسیت 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 به کار برده می‌شوند، معنای عددی و کمی 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 به کار برده می‌شود. در اندازه‌گیری ترتیبی می‌توانیم تعیین کنیم که یک مشاهده (Case) بزرگتر از، مساوی با، یا کوچکتر 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)، (بالای ۲۰ سال، ۱۵ تا ۲۰، ۱۰ تا ۱۵، ۵ تا ۱۰، زیر ۵ سال)، (بالای ۲۰ 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 اصلاً 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 یک 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:** (۱) وجه مشترک سطوح اندازه‌گیری چهارگانه اسمی، رتبه‌ای، فاصله‌ای و نسبی، در طبقه‌بندی آنها است.
- (۲) مقیاس اسمی و ترتیبی (رتبه‌ای) در سطح کیفی‌اند و مقیاس فاصله‌ای و نسبی در سطح کمی‌اند.
- (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.



سوالات چهار گزینه‌ای فصل اول

- ۱- در تحقیقی از آزمودنی‌ها خواسته شده است ردیف تولد خود را در خانواده در محل تعیین شده بنویسند. مقیاس اندازه‌گیری ردیف تولد فرزندان، یک مقیاس است.
- (۱) فاصله‌ای (۲) نسبی (۳) رتبه‌ای (۴) اسمی
- ۲- کدام مقیاس اندازه‌گیری صرفاً به تعیین طبقاتی می‌پردازد که افراد، اشیاء یا حوادث را می‌توان در آنها قرار داد؟
- (۱) اسمی (۲) ترتیبی (۳) فاصله‌ای (۴) نسبی
- ۳- در کدام مقیاس اندازه‌گیری، تمام عملیات آمار و ریاضی را می‌توان انجام داد؟
- (۱) ترتیبی (۲) فاصله‌ای (۳) نسبی (۴) اسمی
- ۴- وقتی که افراد را از نظر تیپ شخصیتی به سه گروه درون‌گرا، متوسط و برون‌گرا تقسیم می‌کنیم، از کدام مقیاس اندازه‌گیری استفاده کرده‌ایم؟
- (۱) اسمی (۲) ترتیبی (۳) فاصله‌ای (۴) نسبی
- ۵- 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
- ۶- موارد استفاده میانه بیشتر در کدامیک از مقیاس‌های اندازه‌گیری است؟
- (۱) اسمی (۲) رتبه‌ای (۳) فاصله‌ای (۴) نسبی
- ۷- اعدادی که در مقیاس اسمی به کار برده می‌شوند، کدامیک از خصوصیات انتزاعی اعداد را دارا می‌باشند؟
- (۱) مشخص بودن (۲) مرتب بودن (۳) ترکیبی بودن (۴) مقداری بودن
- 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- کدام مقیاس جهت نام‌گذاری و طبقه‌بندی و ترتیب طبقات به کار می‌رود؟
- (۱) اسمی (۲) نسبی (۳) رتبه‌ای (۴) فاصله‌ای
- ۱۱- کدامیک از گزینه‌های زیر متغیر گسسته است؟
- (۱) تعداد صندلی‌ها (۲) قد مردم (۳) نمره پیشرفت تحصیلی (۴) وزن دانشجویان
- ۱۲- مناسب‌ترین مورد استفاده از نما زمانی است که مقیاس اندازه‌گیری به کار رفته مقیاس باشد.
- (۱) اسمی (۲) ترتیبی (۳) فاصله‌ای (۴) نسبی
- ۱۳- نمونه‌ای ۸۱ نفری انتخاب نموده‌ایم. در صورتی که میزان انحراف معیار جامعه $4/5$ باشد میزان خطای معیار میانگین چقدر است؟
- (۱) $0/50$ (۲) ۱ (۳) $1/5$ (۴) ۲
- ۱۴- با کدام روش آمار 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
- ۱۵- هدف از کاربرد آمار استنباطی چیست؟
- (۱) تعمیم یافته‌ها به جامعه‌ی مرجع (۲) مشخص نمودن میزان خطای شاخص‌ها
(۳) تفسیر آماره‌های محاسبه شده (۴) محاسبه شاخص‌ها یا آماره‌ها

پاسفنامه سوالات چهارگزینه‌ای فصل اول

- ۱- گزینه ۳ صحیح است.
مقیاس اندازه‌گیری ردیف تولد فرزندان، یک مقیاس رتبه‌ای است. به کلمه‌ی ردیف (ترتیب) توجه شود.
- ۲- گزینه ۱ صحیح است.
مقیاس اندازه‌گیری اسمی صرفاً به تعیین طبقاتی می‌پردازد که افراد، اشیاء یا حوادث را می‌توان در آنها قرار داد.
- ۳- گزینه ۳ صحیح است.
در مقیاس اندازه‌گیری نسبی، تمام عملیات آماری و ریاضی را می‌توان انجام داد.
- ۴- گزینه ۱ صحیح است.
وقتی که افراد را از نظر تیپ شخصیتی به سه گروه درون‌گرا، متوسط و برون‌گرا تقسیم می‌کنیم، از مقیاس اندازه‌گیری اسمی استفاده کرده‌ایم.
- ۵- گزینه ۳ صحیح است.
مقیاس اندازه‌گیری دو متغیر «وزن» و «بهره هوشی» به ترتیب عبارتند از: نسبی - فاصله‌ای.
- ۶- گزینه ۲ صحیح است.
میان‌بیشتر در مقیاس اندازه‌گیری رتبه‌ای به کار می‌رود.
- ۷- گزینه ۱ صحیح است.
در مقیاس اسمی اعداد مورد استفاده فقط مشخص هستند.
- ۸- گزینه ۴ صحیح است.
میزان زاد و ولد در شهر تهران گسسته است، چون اعشاری‌پذیر نیست.
- ۹- گزینه ۴ صحیح است.
جنسیت یک متغیر ناپیوسته است.
- ۱۰- گزینه ۳ صحیح است.
مقیاس رتبه‌ای جهت نام‌گذاری و طبقه‌بندی و ترتیب طبقات به کار می‌رود.
به کلمه‌ی ترتیب توجه شود.
- ۱۱- گزینه ۱ صحیح است.
تعداد صندلی‌ها متغیر گسسته است.
- ۱۲- گزینه ۱ صحیح است.
مناسب‌ترین مورد استفاده از نما زمانی است که مقیاس اندازه‌گیری به کار رفته مقیاس اسمی باشد.

شما	شاخص اسمی
میان	شاخص ترتیبی
میانگین	شاخص فاصله‌ای

۱۳- گزینه ۱ صحیح است.

$$\text{خطای معیار میانگین} = \frac{\text{انحراف معیار جامعه}}{\sqrt{\text{نمونه}}} = \frac{4/5}{9} = 0/50$$

۱۴- گزینه ۳ صحیح است.

با استفاده از روش آماری توصیفی می‌توان افزایش تولید ناخالص ملی ایران را در سه سال گذشته نمایش داد.

۱۵- گزینه ۱ صحیح است.

هدف آمار استنباطی تفسیر، قضاوت و استنباط و در نهایت تعمیم یافته‌ها به جامعه است.

بخش دوم

آمار توصیفی و کارکردهای آن

عناوین اصلی

❖ فصل دوم: سازمان دهی و نمایش داده‌ها (نمودارها)

❖ فصل سوم: نشان دادن مرکزیت داده‌ها

(پارامترهای مکان مرکزی)

❖ فصل چهارم: پارامترهای پراکندگی

فصل دوم

سازمان‌دهی و نمایش داده‌ها (نمودارها)

عناوین اصلی

❖ سازمان‌دهی داده‌ها

❖ نمایش داده‌ها (نمودارها)

فصل دوم

سازمان دهی و نمایش داده‌ها (نمودارها)

سازماندهی داده‌ها

با توجه به اینکه اطلاعات بدست 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، کمی 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 یک 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. یک 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 نزدیک به ۱۰، ولی اگر توزیع شامل تعداد زیادی نمره است، برای مثال بیش از ۵۰۰، آنگاه طبقات را نزدیک به ۲۰ انتخاب می‌نماییم. در گام دوم فاصله بزرگترین و کوچکترین 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 $4/5$ 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 بدست آوریم.

حدود واقعی عدد ۱۹ ← $19/5 - 18/5$

حدود واقعی عدد $19/5$ ← $19/5 - 18/5$

حدود واقعی عدد $12/25$ ← $12/25 - 11/25$

در مثال بالا طبقه‌بندی پیوسته با احتساب حدود واقعی اعداد، اینگونه می‌شود:

- $1/5 - 4/5$
- $4/5 - 7/5$
- $7/5 - 10/5$
- $10/5 - 13/5$



مراحل طبقه‌بندی:

۱- مرتب‌کردن داده‌ها از کوچک به بزرگ

۲- بدست 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

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

۴- تعیین فاصله طبقات.

برای تعیین فاصله طبقات باید دامنه تغییرات را تقسیم بر تعداد طبقات کرد:

$$I = \frac{R(\text{دامنه طبقات})}{K(\text{تعداد طبقات})}$$

۵- 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 به-

دست می‌آید:

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

مثلاً در طبقه ۴/۵-۱/۵ حد میانی طبقه عدد ۳ است:

$$\frac{۱/۵ + ۴/۵}{۲} = ۳$$

مثال: در زیر نمرات درس جامعه‌شناسی یک کلاس آورده شده است. توزیع فراوانی این نمرات را با فاصله طبقاتی مناسب به دست آورید.

نمرات	۱۸	۱۵	۲۰	۱۲	۱۴	۱۸/۵	۱۹	۱۳	۱۷
	۱۲/۲۵	۱۶	۱۹/۵	۱۵/۷۵	۱۱	۹	۲۰	۱۸/۲۵	۱۶/۷۵

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

$$R = ۲۰ - ۹ + ۱ = ۱۲$$

۴ = تعداد طبقات

$$\text{فاصله طبقات} = i = \frac{۱۲}{۴} = ۳$$

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

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

جدول توزیع فرآوانی نمرات درس جامعه شناسی گروهی از دانش آموزان

نکته: فرآوانی تراکمی درصدی از این فرمول بدست می‌آید: برای این کار فرآوانی تراکمی هر طبقه $\times 100 \times \frac{CF}{N}$ 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 فرآوانی بدست می‌دهد. هدف نمودار، ساده‌سازی است. نمودار به تفهیم مطلب کمک کرده و توجه را معطوف به ماهیت 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

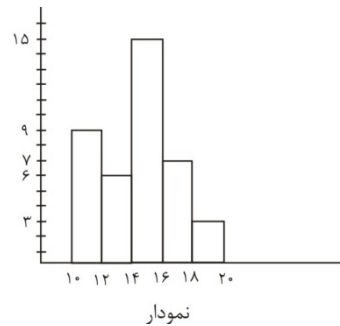
هیستوگرام نموداری است که در آن، فرآوانی‌ها به صورت سطوحی عمودی نشان داده می‌شوند. این نمودار ویژه **متغیرهای کمی (فاصله‌ای و نسبی) و همچنین داده‌های پیوسته** است. این نمودار از یک سری ستون‌های به هم چسبیده تشکیل می‌شود، که در آن هر ستون نشان‌دهنده طبقه‌ای از اعداد است؛ عرض هر ستون برابر فاصله طبقه و ارتفاع آن مساوی فرآوانی 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: اگر نمره ریاضی ۲۰ نفره دانشجوی به شکل زیر دسته‌بندی شده باشد:

فرآوانی مطلق	طبقات نمره
۹	۱۰ تا ۱۲
۶	۱۲ تا ۱۴
۱۵	۱۴ تا ۱۶
۷	۱۶ تا ۱۸
۳	۱۸ تا ۲۰

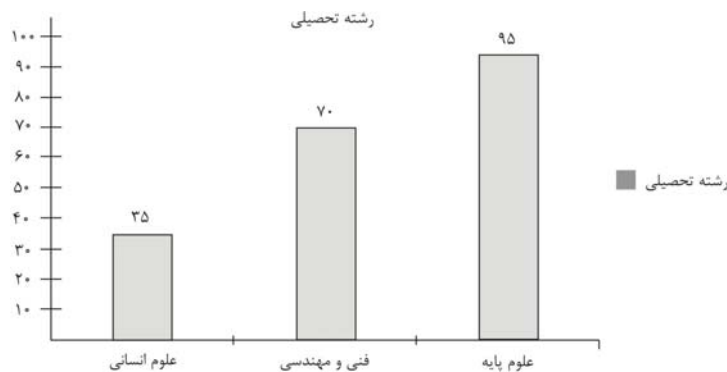


نمودار هیستوگرام آن به شکل زیر خواهد بود:



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 جدائی از یکدیگر هستند، به آسانی مورد استفاده قرار داد. این نمودار ویژه‌ی متغیرهای اسمی و رتبه‌ای است. در این نمودار باید داده‌ها گسسته باشند.

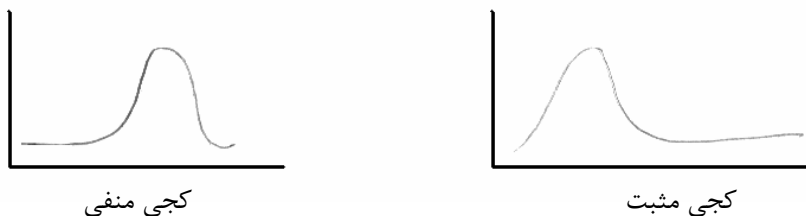


۳- نمودار چندضلعی فراوانی (نمودار چندبر یا پلی‌گون)

چنانچه در یک نمودار مستطیلی (هیستوگرام) نقطه‌ی وسط طبقات مختلف را از طریق یک 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 مناسب‌تر است. این نمودار برای متغیرهای کمی (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.

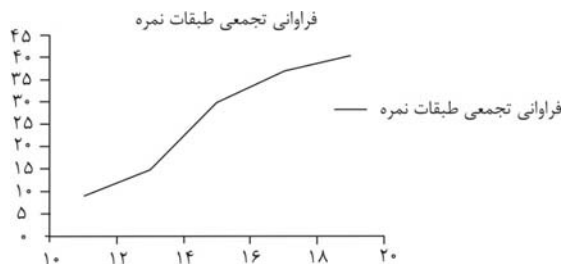


نمودار متقارن: نمودار چندضلعی وقتی متقارن است که نیمی از آن عیناً همانند نیمی دیگر باشد. هنگامی که یک نیمه، عیناً همانند نیمه دیگر نباشد، نمودار چندضلعی نامتقارن نامیده می‌شود؛ که این نمودار خود به دو دسته تقسیم می‌شوند؛ هنگامی که دنباله راست منحنی طولانی‌تر از دنباله چپ آن باشد منحنی را دارای کجی به 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 اُجیو (Ogive)

اگر مبنای منحنی خط بر اساس فراوانی تجمعی باشد نمودار حاصله اُجیو 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 ۷۵ درصد 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:

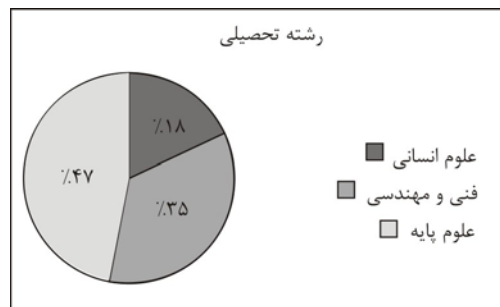
$$S_i = \frac{F_i}{N} \times 100^\circ$$

ترسیم نمودار دایره‌ای بر حسب هر یک از قطعاتها و نیز سهم هر قطعه در کل دایره را می‌توان در قالب درصد نشان داد. به عنوان مثال اگر توزیع فراوانی رشته‌های تحصیلی مختلف در یک 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 چارکها و حداقل و حداکثر مشاهدات است. بدین ترتیب که جعبه مشاهدات اختلاف چارک اول و سوم است. در این نمودار ابتدای جعبه چارک اول و انتهای آن چارک سوم است. خطی که جعبه را به دو قسمت تقسیم می‌کند میانه‌ی مشاهدات است.

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

پراکندگی به میزان تجمع 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 نمره‌ها در سمت چپ منحنی قرار می‌گیرند و هرگاه آزمونی را ساده برگزار کنیم کجی منفی خواهد بود زیرا اکثریت نمره‌ها در سمت راست منحنی قرار خواهند گرفت.

سوالات چهار گزینه‌ای فصل دوم

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

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

۲- در یک آزمون کلاسی ۷۵ درصد نمره قبولی، ۱۵ درصد نمره تجدیدی و ۵ درصد نمره مردودی گرفته‌اند. برای نشان دادن وضعیت تحصیل در این کلاس چه نموداری از همه مناسب‌تر است؟

- (۱) چندضلعی (۲) دایره‌ای (۳) ستونی (۴) هیستوگرام

۳- 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)

۴- نمایش هندسی نتایج آزمون‌ها را می‌نامند؟

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

۵- کدامیک از انواع نمودارها برای نشان دادن رابطه اجزا با کل داده‌ها به کار می‌رود؟

- (۱) دایره‌ای (۲) ستونی (۳) چندضلعی (۴) چندضلعی تراکمی

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

- (۱) میلله‌ای (۲) هیستوگرام (۳) ستونی (۴) فراوانی تجمعی (آجایو)

۷- 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

۸- مناسب‌ترین نمودار جهت داده‌های ناپیوسته چه نموداری است؟

- (۱) چند ضلعی فراوانی (۲) چند ضلعی تراکمی (۳) هیستوگرام (۴) میلله‌ای

۹- 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

۱۰- در نمودار هیستوگرام عرض هر ستون معرف چیست؟

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

۱۱- کدامیک از نمودارهای زیر برای ترسیم داده‌های گسسته به کار می‌رود؟

- (۱) چند ضلعی (۲) چند ضلعی تراکمی (۳) ستونی (۴) هیستوگرام

۱۲- 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