مجله دیابت و لیپید ایران. دو ماهنامه آذر- دی ۱۳۹۱؛ دوره ۱۲ (شماره ۲) ۱۰۰-۸۳

سری آمار: خلاصه کردن و نمایش داده‌ها

محمد اصغری جعفرآبادی[1*]، سیده مومنه محمدی[2]

چکیده

ارائه نتایج آماری و تنظیم بخش یافته‌ها در گزارش مطالعات، یکی از دغدغه‌های پژوهشگران علوم پزشکی است. هدف از این مقاله معرفی روش‌های ساده و کارا برای خلاصه کردن و نمایش دادن انواع داده‌های حاصل از مطالعات علوم پزشکی بود. تعاریف اولیه انواع متغیرها و روش‌های خلاصه کردن و ارائه کردن آنها در حالت‌های مختلف معرفی گردید.برای خلاصه کردن متغیرهای کیفی و کمی به ترتیب فراوانی (درصد) و میانگین (انحراف معیار) معرفی گردید و برای برخی موارد استثناء شاخص میانه (صدک ۲۵ – صد ۷۵) پیشنهاد گردید. برای وضعیت‌های چند متغیره نحوه تعمیم محاسبات ارائه گردید. همچنین برای نمایش داده‌ها، روش نموداری و جدولی متناسب با ماهیت آنها در حالت‌های یک و چند متغیره توصیه شد.

واژگان کلیدی: آمار توصیفی، خلاصه کردن، نمایش، متغیر، کیفی، کمی

۱- مرکز تحقیقات آموزش علوم پزشکی، گروه آمار و اپیدمیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی تبریز، تبریز، ایران
۲- گروه علوم تشریحی، دانشکده پزشکی، دانشگاه علوم پزشکی تبریز، تبریز، ایران

***نشانی**: تبریز، خیابان گلگشت، خیابان عطار نیشابوری، دانشکده بهداشت، دانشگاه علوم پزشکی تبریز، گروه آمار و اپیدمیولوژی، کدپستی: ۵۱۶۶۶۱۴۷۱۱، تلفن: ۲- ۰۴۱۱۳۳۵۷۵۸۰، نمابر: ۰۴۱۱۳۳۴۰۶۳۴، پست الکترونیک: asgharimo@tbzmed.ac.ir

تاریخ دریافت: ۱۳۹۱/۸/۱۰ **تاریخ درخواست اصلاح: ۱۳۹۱/۰۸/۲۰** **تاریخ پذیرش: ۱۳۹۱/۰۸/۲۷**

[Downloaded from journals.tums.ac.ir on 2024-11-16]

مقدمه

آمار علم تحلیل مطالعات کمی است. روش‌های آماری مجموعه‌ای از ابزار برای توصیف و تحلیل داده‌ها شامل بررسی روابط در داده‌ها و ارائه کردن آنها می‌باشد که به ترتیب با عنوان آمار توصیفی و تحلیلی (استنباطی) نامیده می‌شوند. یکی از جنبه‌های بسیار مهم در این زمینه، نحوه خلاصه کردن و ارائه نمودن داده‌هاست. انتخاب مناسب روش خلاصه‌سازی و گزارش کردن داده‌ها می‌تواند یک محتوی و مطلب را به شکلی متفاوت جلوه دهد. از این رو نیاز است در مورد خلاصه‌سازی و ارائه کردن داده‌ها تکنیک‌هایی فرا گرفته شود تا این امر به بهترین نحو ممکن صورت پذیرد. لازم به ذکر است که هر نوع داده‌ای تکنیک خاصی برای ارائه کردن نیاز دارد به عنوان مثال در بخش یافته‌های یک مقاله عبارت زیر را می‌توان یافت، که از تعداد کل ۲۰۰ نفر شرکت کننده در مطالعه ۵۰ نفر (۲۵ درصد) زن بودند، میانگین سنی افراد حاضر در مطالعه ۳۶ سال (انحراف معیار ۵/۴) بود و میانه تعداد مراجعه به مرکز بهداشت در این افراد ۵ (حداقل ۱ و حداکثر ۸) بار بوده است. همان‌طور که ملاحظه شد، نحوه ارائه کردن داده‌ها به نوع آنها بستگی دارد، بنابراین ابتدا نیاز است کلیاتی در مورد نوع داده‌ها ارائه گردد.

متغیرها و داده‌ها

همان‌طور که برای ساختن یک ساختمان، مصالح اولیه‌ای شامل بلوک سیمان آهن ماسه و ... نیاز است، برای انجام تحلیل‌های آماری نیز مصالح اولیه‌ای نیاز است که شامل متغیرها و داده‌ها هستند. متغیرها به صفاتی گفته می‌شوند که هدف اندازه‌گیری و ارزیابی می‌باشند؛ نظیر قد، وزن، نمایه توده بدنی، جنسیت، سطح تحصیلات و ... داده‌ها مشاهدات عددی حاصل از ارزیابی هستند که در قالب اعداد ظاهر می‌شوند؛ مثلاً در اندازه‌گیری نمایه توده بدنی، داده‌ها شامل ۲۲، ۱۹، ۳۲، ۲۸ و ... (بر حسب کیلوگرم بر متر مربع) می‌باشند. در واقع متغیر به صفتی گفته می‌شود که قابلیت ارزش‌گذاری (انتساب عدد) داشته باشد و از مشاهده‌ای به مشاهده‌ی دیگر تغییر کند. داده‌ها مشاهدات عددی حاصل از این ارزش‌گذاری و تغییرپذیری هستند که ورودی‌های اولیه تحلیل‌های آماری را می‌سازند و ممکن است حاصل از ارزش‌گذاری مربوط به صفاتی نظیر جنسیت و یا ممکن است ناشی از اندازه‌گیری‌های فیزیکی نظیر وزن باشند [۱]. در مقابل متغیر مفهوم ثابت وجود دارد، مثلاً اگر در مطالعه‌ای آزمودنی‌ها فقط خانم‌ها باشند، جنسیت دیگر یک متغیر نیست [۲].

مقیاس‌های اندازه‌گیری متغیرها

همان طور که اشاره شد، آشنایی با اصول ارزش‌گذاری داده‌ها خصوصیات و مزایا و معایب هر یک از آنها درک صحیحی در انتخاب روش آماری مناسب برای توصیف و تحلیل مناسب با نوع داده‌ها فراهم می‌کند. برای ارزش‌گذاری صفات از یکی از سه روش شمارش مرتب کردن و طبقه‌بندی کردن استفاده می‌شود:

۱- برای اندازه‌گیری قد بیماران، متر به کار می‌رود و تعداد سانتی‌مترهای موجود در روی متر **شمارش** می‌شود.

۲- برای ارزیابی سطح تحصیلات این سطح با کدهای ۱ تا ۴ به ترتیب برای سطوح "بی‌سواد" تا "تحصیلات دانشگاهی" **مرتب** می‌شوند.

۳- برای بررسی بروز یک بیماری افراد پس از تشخیص در یکی از دو دسته "سالم" یا "بیمار" **طبقه‌بندی** می‌شوند [۱].

بر اساس شیوه ارزیابی ارائه شده فوق مقیاس‌های اندازه‌گیری تعریف می‌شوند.

اندازه‌گیری بر اساس شمارش

همان طور که اشاره شد در این روش متناسب با ماهیت داده‌ها تعداد واحدهای اندازه‌گیری تشکیل دهنده شمارش می‌شوند تا اندازه‌گیری مورد نظر را نتیجه دهند.

مثلاً برای اندازه‌گیری قد بیمار تعداد سانتی‌مترهای روی متر شمارش می‌شود، برای اندازه‌گیری دمای هوا تعداد درجات سلیوس یا فارنهایت روی دماسنج شمارش می‌شود و برای بررسی تعداد افراد مبتلا به سرطان کولورکتال افراد مبتلا به این بیماری شمارش می‌شوند. در

[Downloaded from journals.tums.ac.ir on 2024-11-16]

هر یک از این موارد ارزش‌گذاری انجام شده به یک عدد واقعی منجر می‌شود که به همین دلیل به این ارزش‌گذاری اندازه‌گیری استاندارد گفته می‌شود و متغیرهای حاصل از این اندازه‌گیری‌ها، متغیرهای کمی نامیده می‌شوند.

در مثال سوم به طور مطلق و شامل شمارش تعداد افراد می‌باشد که بر حسب مجموعه‌ی اعداد حسابی انجام می‌شود، در حالی که در شمارش‌های مثال‌های ۱ و ۲ شمارش یک " اندازه" یا "بزرگی" است و از این جهت نسبی یا قراردادی می‌باشد (اندازه‌گیری‌هایی نظیر قد، وزن، پهنا، دما، سرعت و... از این نوع می‌باشند). در اندازه‌گیری قد بیمار و دمای هوا، فاصله بین دو عدد متوالی در متر و دماسنج یکسان است، یعنی اندازه‌گیری‌های فوق متساوی‌الفاصله هستند. این خصوصیت سبب می‌شود که عملگرهای معمولی حسابی را (نظیر جمع تفریق ضرب تقسیم) در مورد آنها به کار برد که در نهایت بر اساس آنها می‌توان شاخص‌های مناسب آماری برای توصیف داده‌ها را محاسبه نمود (نظیر میانگین ساده[1]).

اندازه‌گیری‌های گسسته و پیوسته

بین اندازه‌گیری هوا و قد بیماران با شمارش تعداد بیماران یک تفاوت وجود دارد و آن هم این است که دو اندازه‌گیری اول در یک طیف پیوسته صورت می‌گیرد و بین هر دو عدد متوالی تعریف شده می‌توان عدد سومی را یافت که از لحاظ اندازه‌گیری مورد نظر معنی دارد. حال آنکه شمارش تعداد بیماران دقیقا روی اعداد طبیعی قرار دارد و مثلاً ۰، ۱، ۲، ۳ و... و ۱۰ بیمار می‌توان شمرد و مثلاً ۶/۳۸۴ بیمار معنی ندارد. به اندازه‌گیری‌های از نوع ول پیوسته و از نوع دوم پیوسته گفته می‌شود.

مقیاس‌های نسبتی و فاصله‌ای

در مثال‌های فوق بین شمارش تعداد بیماران و اندازه‌گیری‌های قد آنها با دمای هوا یک تفاوت اساسی و ساختاری وجود دارد؛ در مورد دو اندازه‌گیری شمارش افراد و اندازه‌گیری قد صفر معنی دارد صفر نفر، صفر متر یا سانتی‌متر که نشان می‌دهد واقعاً صفت مورد نظر وجود ندارد (مثلاً فردی یا طولی وجود ندارد)، حال آنکه در مورد دمای هوا بر حسب (سانتی‌گراد یا فارنهایت) صفر یک نقطه قراردادی است و به معنای عدم هر گونه دما نیست، چون حتی در نقطه پایین‌تر از این نقطه مثلاً در ۱۰- درجه سانتی‌گراد هم دمایی وجود دارد. اندازه‌گیری‌هایی که دو خصوصیت متساوی‌الفاصله و صفر مطلق را داشته باشند، نسبتی نامیده می‌شوند به عبارت دیگر وجود این دو خصوصیت سبب می‌شود که در این اندازه‌گیری‌ها، چند برابر بودن (نسبت) معنی داشته باشد؛ مثلاً ۲ متر دو برابر ۱ متر است.

اندازه‌گیری بر اساس مرتب کردن

زمانی که دمای بدن دو مریض به ترتیب ۳۷ و ۳۹ درجه سانتی‌گراد است، می‌توان گفت که دمای بدن مریض دوم بالاتر از مریض اول است که به نوعی برتری یا ترجیح در اندازه‌ها را نشان می‌دهند و به همین دلیل به آنها ترتیبی گفته می‌شود، چون آنها ترتیبی از اندازه، کمیت و یا بزرگی را نشان می‌دهند. در برخی ارزیابی‌های کلینیکی امکان ارزیابی کمی با مقیاس متساوی‌الفاصله وجود نداشته باشد، مثلاً برای تعیین مرحله سرطان بر اساس دسته‌بندی AJCC، در چهار مرحله I، II، III و IV یک ابزار فیزیکی نظیر ترازو، متر یا سنجه‌ای وجود ندارد که بتواند به طور متساوی‌الفاصله، متغیر مرحله سرطان را درجه‌بندی کند. این موضوع در اغلب ارزیابی‌های بالینی که به ترجیح، نگرش عقیده یا قضاوت‌های بالینی بر می‌گردد، وجود دارد. در صورتی که برای اندازه‌گیری یک موضوع دو انتخاب مقیاس متساوی‌الفاصله یا ترتیبی وجود داشته باشد، در این صورت انتخاب بهتر، مقیاس متساوی‌الفاصله است، زیرا اطلاع بیشتری در مقایسه با ترتیبی دارد.

مقیاس‌های رتبه‌ای

مقیاس‌های رتبه‌ای به دو صورت قابل دسته‌بندی هستند: مقیاس‌های رتبه‌ای مبتنی بر اولویت‌بندی و مقیاس‌های رتبه‌ای مبتنی بر اختصاص رتبه. به عنوان مثال، از یک پزشک متخصص گوارش و بیماری‌های کبد خواسته

۱- جمع مشاهدات تقسیم بر تعداد داده‌ها

[Downloaded from journals.tums.ac.ir on 2024-11-16]

می‌شود که درمان‌های جراحی، شیمی‌درمانی، بایوپسی و دارویی را برای یک بیمار سرطانی با شرح مشخص اولویت‌بندی کند، در این صورت ممکن است نظر این پزشک به صورت زیر باشد:

۱) جراحی، ۲) بایوپسی، ۳) دارویی و ۴) شیمی‌درمانی

در اولویت‌بندی فوق، اندازه تفاوت بین درمان‌های شیمی‌درمانی و جراحی، جراحی و دارویی، یا دارویی با بایوپسی اهمیتی ندارد و تنها در اولویت بودن آنهاست که اهمیت می‌یابد. نداشتن خصوصیت متساوی‌الفاصله در این مقیاس سبب می‌شود که عملگرهای ریاضی نظیر جمع و تفریق در مورد این مقیاس قابل اعمال کردن نباشد و در نتیجه نتوان میانگین را برای چنین داده‌هایی محاسبه کرد.

مقیاس‌های رتبه‌ای اطلاع بیشتری در مقایسه با مقیاس‌های مبتنی براولویت‌بندی فراهم می‌کنند. برای روشن شدن موضوع، در مثال ارائه شده در بخش فوق، فرض کنید هر یک از روش‌های درمانی بر اساس نظر پزشک یکی از گزینه ۱ تا ۵ را که به صورت زیر تعریف می‌شوند، اختصاص دهند: ۱) کاملاً مخالف، ۲) مخالف، ۳) بی نظر، ۴) موافق و ۵) کاملاً موافق

اختصاص عدد به ترتیب پاسخ‌ها به آن خاصیتی می‌بخشد که فراتر از رتبه‌دهی بر مبتنی بر اولویت‌بندی عمل می‌کند، به طوری که حتی در برخی از موارد مخصوصاً زمانی که طیف رتبه‌ها بیشتر می‌شود (نظیر ارزیابی درد بر اساس مقیاس آنالوگ بصری VAS)، می‌توان آن را با مقیاس فاصله‌ای قابل مقایسه دانست، که به نوعی آن را خاصیت فازی این مقیاس دانسته‌اند [۳].

اندازه‌گیری‌های مبتنی بر مرتب کردن آیتم‌ها: ارزیابی‌های رسته‌ای (متغیرهای کیفی)

در این ارزیابی‌ها، اگر خصوصیات مجموعه‌ای از آیتم‌ها با ملاک خاصی همخوانی داشته باشد به آن دسته اختصاص می‌یابد و در غیر این صورت به دسته دیگری اختصاص داده می‌شود. به عنوان مثال، دسته‌بندی عده‌ای از افراد بر حسب جنسیت که به دو گروه «زن» و «مرد» اختصاص داده می‌شوند و یا دسته بندی عده‌ای از افراد بر حسب سطح تحصیلات که به گروه‌های «دیپلم و زیردیپلم»، «لیسانس»، «فوق لیسانس» و «دکتری» اختصاص می‌یابند. همان‌طور که ملاحظه می‌شود، بین دو مثال فوق یک تفاوت وجود دارد؛ در دسته‌بندی جنسیت به گروه‌های زن و مرد هیچ‌گونه برتری در رده‌ها وجود ندارد، در حالی که در دسته بندی مثال دوم می‌توان بیان نمود که سطح تحصیلات «دکتری» نسبت به «دیپلم و زیر دیپلم» بالاتر است. به عبارت دیگر نوعی برتری و ترجیح در سطح تحصیلات را می‌توان ملاحظه نمود. متغیرهایی نظیر جنسیت (و نژاد، مذهب، رنگ چشم، گروه های خونی و...) اصطلاحاً اسمی (Nominal) نامیده می‌شوند و کدها صرفاً برای شناسایی است و جنبه برتری در آنها مشاهده نمی‌شود. متغیرهایی نظیر سطح تحصیلات اصلاحاً ترتیبی (Ordinal) نامیده می‌شوند و کدهایی که به آنها اختصاص داده می‌شود (مثلاً ۱، ۲، ۳ و ۴ به ترتیب برای «دیپلم و زیر دیپلم»، «لیسانس»، «فوق لیسانس» و «دکتری»)، به نوعی برتری را نشان می‌دهد. با توجه به خصوصیتی که این متغیرها دارند، به طوری که بر اساس دسته‌بندی کردن آنها به گروه‌ها، این متغیرها ارزیابی می‌شوند، **نمی‌توان** در مورد آنها اعمال جمع، تفریق، ضرب و تقسیم و یا میانگین‌گیری را انجام داد، اما در ارزیابی‌های مرتبط با این متغیرها معمولاً تعداد موارد موجود در هر رده شمرده می‌شود و برای نمایش این متغیرها به ازای رده‌های تعریف شده در متغیر، فراوانی و درصد مشاهدات گزارش می‌شود.

نکات اساسی برای چهار مقیاس

تنها در مقیاس نسبتی صفر مطلق وجود دارد و در سایر مقیاس‌ها صفر قراردادی است. هر مقیاسی خواص مقیاس قبلی را دارد و بنابراین بهترین نوع مقیاس برای اندازه‌گیری، مقیاس نسبتی است، چون خواص سایر مقیاس‌ها نیز در بر دارد. روش‌های آماری برای داده‌های مقیاس‌های سطح پایین را می‌توان (با اندکی چشم‌پوشی) برای داده‌های با مقیاس بالاتر به‌کار برد ولی عکس حالت امکان‌پذیر نیست. اندازه‌گیری‌ها بر اساس هدف مطالعه؛ درآمد به صورت رتبه‌ای (سطح درآمد) و یا نسبتی (میزان درآمد)، انجام می‌شود.

[Downloaded from journals.tums.ac.ir on 2024-11-16]

خلاصه‌سازی و نمایش داده‌ها برای متغیرهای کیفی:

خلاصه‌سازی داده‌ها برای متغیرهای کیفی

این مجموعه از متغیرها شامل متغیرهای اسمی و رتبه‌ای می‌باشند. برای سادگی ابتدا نحوه خلاصه‌سازی و نمایش داده‌های اسمی دو حالتی ارائه می‌شود و سپس نتایج برای متغیرهای اسمی بیش از دو حالتی و رتبه‌ای تعمیم داده می‌شود.

در ارزیابی تعداد مرگ و میر ناشی از سرطان‌های کولون و رکتال به ترتیب تعداد ۲۲۳ و ۱۲۱ مورد مرگ ناشی از این دو سرطان مشاهده شد [۴]. در این مورد علاوه خلاصه کردن این متغیر با استفاده از فراوانی‌های شمرده شده، نسبت آنها به کل نمونه‌ها (برای ایجاد قابلیت مقایسه) به دست می‌آید که شاخص جدیدی به نام سهم (Proportion) را تعریف می‌نماید؛ مثلاً سهم مرگ ناشی از سرطان‌های کولون و رکتال در این مطالعه به ترتیب برابر ۰/۲۷۳ (۸۱۷÷۲۲۳) و ۰/۳۰۱ (۴۰۲÷۱۲۱) بوده است. این شاخص معمولاً در عدد صد ضرب شده و نتیجه به درصد بیان می‌شود (مثلاً ۲۷/۳ و ۳۰/۱ درصد).

در واقع سهم حالت خاصی از نسبت (Ratio) است. شاخص نسبت در حالت کلی برابر حاصل تقسیم دو عدد مثلاً (a÷b) است که در ان a و b شمارش یا عدد مربوط به هر ارزیابی می‌توانند باشند (مثلا شاخص BMI حاصل تقسیم وزن (kg) به قد (m) به توان ۲ یک نسبت است). زمانی که در این نسبت صورت کسر بخشی از مخرج کسر باشد، نسبت همان سهم را ایجاد می نماید.

مباحث خاص

در مطالعات اپیدمیولوژیکی، معمولاً نسبت‌ها با مفاهیم دیگری ترکیب شده و تعاریف جدیدی را ایجاد می‌کنند. به عنوان مثال، درصد مرگ ناشی از افراد مبتلا به سرطان‌های کولون و رکتال به ترتیب برابر ۲۷/۳ و ۳۰/۱ درصد بوده است که به ترتیب در یک مدت زمان پیگیری معین با میانه ۲۱/۱ و ۲۰/۲۷ ماهه، بررسی شده بود. زمانی که نسبت در دوره زمانی معینی مورد بررسی قرار می‌گیرد، اصطلاحاً به آن میزان (Rate) گفته می‌شود. در یک مطالعه‌ی آینده‌نگر[۱] ، اگر در یک زمان معین موارد جدید یک عارضه (پیامد)، نسبت به تعداد کل افراد در معرض خطر این پیامد به دست آید، در این صورت به آن میزان بروز (Incidence Rate) آن پیامد گفته می‌شود، مثلاً میزان بروز سرطان کولون و رکتال در مدت زمان تعریف شده به ترتیب برابر ۲۷/۳ و ۳۰/۱ درصد بوده است. به این شاخص اصطلاحاً مخاطره یا خطر (Risk) نیز گفته می‌شود. در مطالعات گذشته‌نگر[۲] و مقطعی[۳]، مفهوم دیگری به نام شانس مطرح می‌شود که در بخش‌های آتی این سری بدان اشاره خواهد شد. به علاوه در مطالعات بقا[۴] نیز به دلیل ماهیت خاص آنها شاخص دیگری به نام میزان خطر (Hazard Ratio) محاسبه می‌شود که در شماره‌های آتی این سری ارائه خواهد شد.

تعمیم به متغیرهای کیفی چند حالتی

زمانی که مسئله از متغیرهای دو حالتی به متغیرهای بیش از دو حالتی تبدیل می‌شود، با دو شکل از متغیرها روبرو یعنی ۱- متغیرهای اسمی با بیش از دو حالت و ۲- متغیرهای ترتیبی و رتبه‌ای خواهیم شد؛ برای متغیرهای اسمی و ترتیبی با بیش از دو حالت، مشابه متغیرهای دو حالتی، تعداد مشاهدات موجود در هر یک از رده‌های متغیر، شمرده شده و فراوانی مشاهدات موجود در هر رده به دست می‌آید. برای ایجاد قابلیت مقایسه، هر یک از فراوانی‌های به دست آمده به تعداد کل فراوانی‌ها (تعداد نمونه) تقسیم شده و درصدها (Percentages) محاسبه می‌شوند.

به عنوان مثال، برای متغیر سطح تحصیلات به ازای هر یک از رده‌های «بی‌سواد»، «راهنمایی»، «دبیرستان» و «دانشگاهی» تعداد و درصد مشاهدات موجود محاسبه می‌شود [۴]. برای متغیرهای زمینه‌ای نیز در اغلب موارد به ازای هر یک از

۱- یعنی مطالعه‌ای که از نقطه مواجهه (Exposure) یا مداخله (Intervention) شروع می‌شود و پیامد (Outcome) ختم می‌شود.

۲- مطالعاتی که از پیامد شروع می‌شوند و سابقه افراد برای حضور یا عدم مواجهه مورد بررسی قرار می‌گیرد.

۳- مطالعاتی که در مقطعی از زمان با ارزیابی همزمان مواجهه و پیامد صورت می‌گیرد.

۴- مطالعات با ماهیت زمان لازم تا رخداد پیشامد مورد نظر.

[Downloaded from journals.tums.ac.ir on 2024-11-16]

رتبه‌های اختصاص داده شده، محاسبه فراوانی و درصد مشاهدات موجود، مخصوصاً زمانی که تعداد رده‌های متغیر کم باشد، پیشنهاد می‌شود. اما زمانی که طیف تعریف شده برای رتبه‌ها از تنوع بیشتری برخوردار باشد، مثلاً طیف‌هایی با تعداد ۵ و یا تعداد بیشتری از گزینه‌ها، در این صورت در برخی از روش‌های آماری مبتنی بر رتبه‌ها (تحت عنوان روش‌های ناپارامتری که در شماره های آینده این سری در مورد آنها بحث خواهد شد)، میانگین رتبه[1] محاسبه می‌شود. همچنین زمانی که طیف رتبه‌ها کمی وسیع‌تر باشد، مثلاً ۵ یا تعداد بیشتری از رتبه‌ها، در این صورت می‌توان شاخص دیگری نظیر میانه[2] را برای مشاهدات محاسبه کرد [۵]. به عنوان مثال برای اندازه‌گیری مفاهیمی نظیر درد و خستگی از یک مقیاس آنالوگ بصری (Visual Analogue Scale; VAS) استفاده می‌شود و خط کشی است که از صفر تا ۱۰ رتبه بندی شده است و صفر عدم درد و ۱۰ نهایت شدت درد را نشان می‌دهد. به دلیل این که طیف استفاده شده در این متغیر وسیع است، در این صورت شاخص میانه برای گزارش آن استفاده می‌شود [۶].

نمایش داده‌ها برای متغیرهای کیفی

نمایش داده‌ها برای متغیرها از طریق ابزارهایی نظیر جداول و نمودارها انجام می‌شود. برای ارائه بهتر داده‌ها نیاز است جداول و نمودارها ویژگی‌های لازم را داشته باشند؛ جدول باید: خودتوصیف باشد، عنوان واضح داشته باشند، عنوان‌های داخلی واضح و خلاصه باشد، حجم مناسبی از اطلاعات داشته باشد، با متن تناسب داشته باشد، بر اساس اندازه، زمان و ... مرتب شود، آماره‌های خلاصه در سطر یا ستون آن برای انجام مقایسه ارائه شود و در آن تعداد مناسبی ارقام اعشار انتخاب شود. نمودار باید: خودتوصیف باشد، زیرنویس واضح داشته باشد، عنوان‌های محور عمودی، افقی واضح باشد، مقیاس مناسب برای محورها انتخاب شود، ساختار داده‌ها مقطعی یا زمانی باید واضح باشد، برای متغیرهای کمی یا کیفی متناسب با ماهیت آنها به طور مناسب انتخاب شود و برای متغیرهای پیوسته یا گسسته متناسب با ماهیت آنها به طور مناسب انتخاب شود. برای نمایش متغیرهای رسته‌ای (دو یا بیش از دو حالتی) از ابزارهای جداول و نمودارهای میله‌ای و دایره‌ای با نمایش فراوانی و درصد استفاده می‌شود. به عنوان مثال، جدول زیر نمایشی از متغیرهای رسته‌ای فراهم می‌کند:

۱- منظور از میانگین رتبه این است که رتبه‌های اختصاص داده شده با یکدیگر جمع شده و بر تعداد کل موارد تقسیم می‌شوند.

۲- منظور از میانه داده‌ها، عددی است که نصف داده‌ها از آن بزرگتر و نصف داده‌ها از آن کوچکترند.

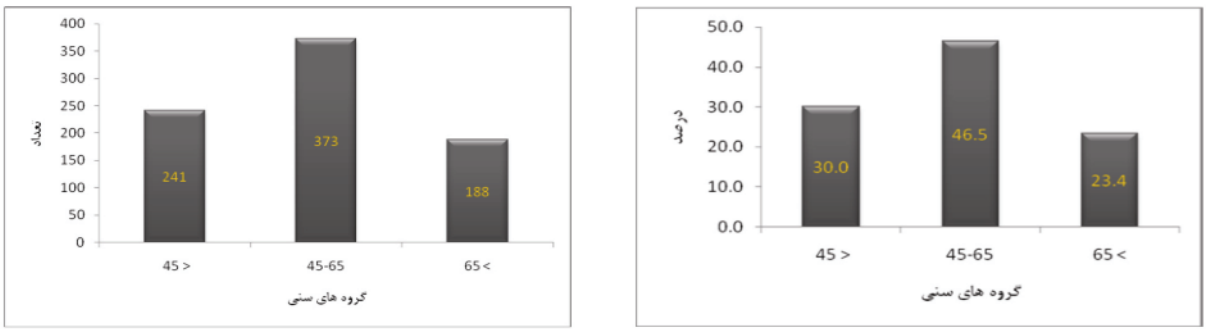
[Downloaded from journals.tums.ac.ir on 2024-11-16]

جدول ۱- فراوانی و درصد افراد مبتلا به سرطان کولون به تفکیک رده‌های متغیرهای سن، جنس و نمایه توده بدنی

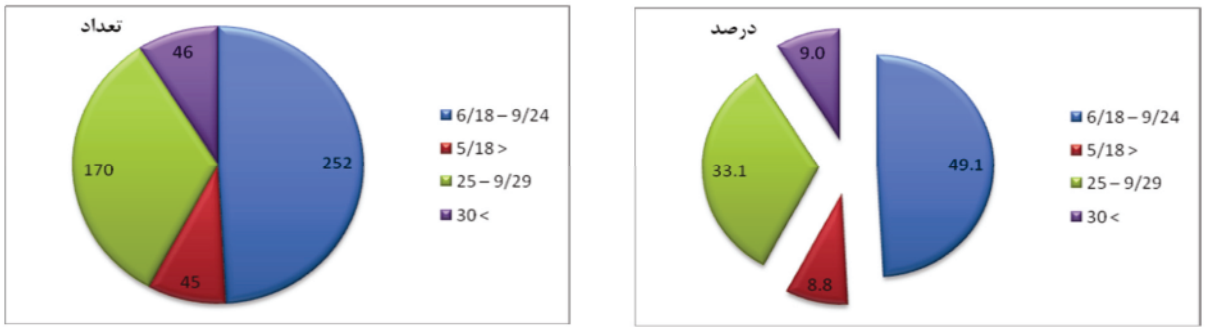
نام متغیر	رده‌های متغیر	فراوانی	درصد
سن حین تشخیص			
	< ۴۵	۲۴۱	۳۰/۰
	۶۵-۴۵	۳۷۳	۴۶/۵
	> ۶۵	۱۸۸	۲۳/۴
جنسیت			
	مرد	۴۷۲	۵۸/۹
	زن	۳۳۰	۴۱/۱
نمایه توده بدنی			
	۱۸/۶ – ۲۴/۹	۲۵۲	۴۹/۱
	< ۱۸/۵	۴۵	۸/۸
	۲۹/۹ – ۲۵	۱۷۰	۳۳/۱
	> ۳۰	۴۶	۹/۰

منبع: [۷]

در جدول ارائه شده، تعداد ارقام مناسب برای اعشار شامل یک رقم می‌باشد. همچنین، نمودارهای ۱ و ۲ نمایش گرافیکی فراوانی (درصد) موارد مبتلا به سرطان کولون مشاهده شده را به ازای هر یک از رده‌های متغیر سن و نمایه توده بدنی، نشان می‌دهند.



نمودار ۱- نمایش فراوانی و درصد افراد مبتلا به سرطان کولون در رده‌های متغیر سن با استفاده از نمودارهای میله‌ای تعداد و درصد



نمودار ۲- نمایش فراوانی و درصد افراد مبتلا به سرطان کولون در رده‌های متغیر شاخص توده بدنی با استفاده از نمودارهای دایره‌ای تعداد و درصد

[Downloaded from journals.tums.ac.ir on 2024-11-16]

همان‌طور که ملاحظه می‌شود، در نمودار میله‌ای هر یک از رده‌های متغیر مورد بررسی و ارتفاع هر میله فراوانی (درصد) مشاهدات موجود در در رده‌ها و در نمودار دایره‌ای هر یک از قطاع‌ها رده‌های متغیر مورد بررسی و میزان درجه هر یک از قطاع‌ها سهم نسبی مشاهدات موجود در رده‌ها از ۳۶۰ درجه (مساحت دایره) را نشان می‌دهد.

برای نمایش متغیرهای رتبه‌ای با طیف وسیع‌تر، نظیر اندازه‌گیری‌های درد با VAS، می‌توان از نمودارهای دیگری نظیر نمودار جعبه‌ای (Box Plot) استفاده کرد.

تعمیم به حالت چند متغیره (Multivariate)

در این بحث، نتایج برای حالت دو متغیره (Bivariate) بررسی شده، تعمیم نتایج آن به حالت چند متغیره سرراست می‌باشد. منظور از خلاصه کردن و نمایش داده‌ها در این حالت، آن است که شاخص‌های تعریف شده فوق به ازای هر یک از رده‌های متغیر اول و به تفکیک رده‌های متغیر دوم ارائه شود. به عنوان مثال، در ارزیابی رابطه سرطان کولورکتال با مصرف سیگار فراوانی مرگ و میر ناشی از این سرطان در جدول ۲ ارائه شده است [۸].

جدول ۲- فراوانی و درصد وضعیت مرگ ومیر افراد مبتلا به سرطان کولورکتال برحسب رده‌های متغیرهای وضعیت کشیدن سیگار

وضعیت کشیدن سیگار	وضعیت مرگ ومیر			
	زنده		مرده	
	فراوانی	درصد	فراوانی	درصد
هرگز مصرف نکرده	۶۰۷	٪۷۳/۰	۲۲۵	٪۲۷/۰
در حال حاضر یا قبلا مصرف کرده	۱۸۷	٪۶۵/۸	۹۷	٪۳۴/۲

بنابرین، همان‌طور که ملاحظه می‌شود، فراوانی (درصد) مرگ ناشی از سرطان کولورکتال (متغیر اول) به ازای هر یک از رده‌های مصرف سیگار (متغیر دوم) محاسبه می‌شود. در جدول فوق درصدها در داخل رده‌های متغیر سطری (وضعیت کشیدن سیگار) محاسبه شده‌اند بنابراین مجموع سطری درصدها برابر ۱۰۰ می‌شود.

مباحث خاص برای حالت چند متغیره

در مطالعات آینده‌نگر، برای متغیرهای رسته‌ای با بیش از دو حالت (مشابه با متغیرهای دو حالتی) شاخص مخاطره به ازای هر یک از رده‌های متغیر مورد بررسی محاسبه می‌شود. به عنوان مثال میزان خطر مرگ ناشی از سرطان کولورکتال در هر یک از رده‌های نوع درمان در یک مدت زمان پیگیری معین با میانه ۲۱/۱ و ۲۰/۲۷ ماهه به ترتیب برابر ۲۸/۲ ٪، ۴۰/۴٪ و ۳۴/۳٪ برای انواع درمان جراحی، شیمی درمانی و بایوپسی بوده است (جدول ۳).

جدول ۳- میزان خطر مرگ افراد مبتلا به سرطان کولورکتال برحسب اولین درمان استفاده شده

اولین درمان استفاده شده	میزان مرگ ومیر ناشی از سرطان کولورکتال	
	فراوانی کل/ فراوانی مرگ و میر	میزان خطر (٪)
جراحی	۲۴۰/۸۵۲	٪۲۸/۲
شیمی‌درمانی، رادیوتراپی و ایمونوتراپی	۴۶/۱۱۴	٪۴۰/۴
بایوپسی	۴۶/۱۳۴	٪۳۴/۳

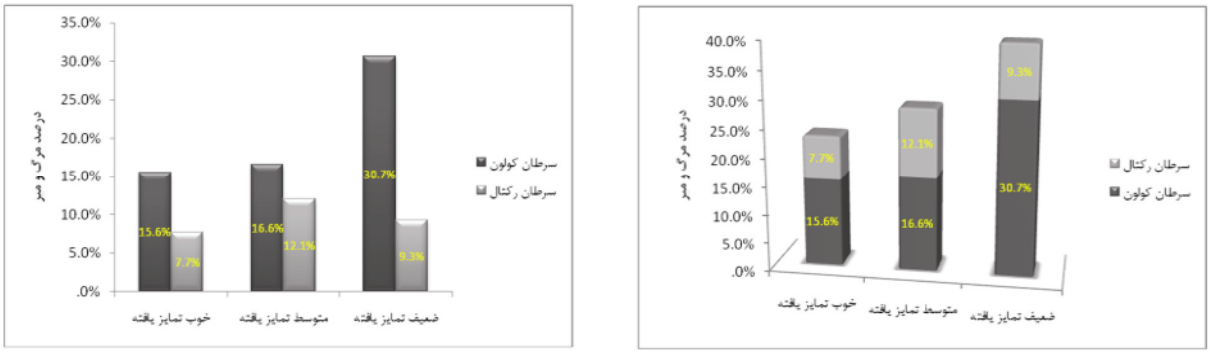
[Downloaded from journals.tums.ac.ir on 2024-11-16]

شایان ذکر است در این شکل از مطالعات برای متغیرهای رتبه‌ای و ترتیبی می‌توان وجود یک روند در مخاطره‌های محاسبه شده را بررسی کرد؛ به عنوان مثال، میزان مخاطره مرگ ناشی از سرطان کولون در مراحل ۱، ۲، ۳ و ۴ سرطان به ترتیب برابر ۱۹/۰٪، ۲۱/۰٪ ، ۲۸/۲٪ و ۵۶/۱ به ترتیب برای مراحل I تا IV سرطان کولورکتال به دست آمد (جدول ۴) که یک روند افزایش در مخاطره مرگ ناشی از سرطان کولون را با افزایش رتبه‌های متغیر مرحله پاتولوژیک تومور، نشان می‌دهد. همچنین در مورد مطالعات مقطعی و گذشته‌نگر نیز می‌توان شانس را به ازای هر یک از رده‌های متغیر مورد بررسی، محاسبه نمود.

جدول ۴- میزان خطر مرگ افراد مبتلا به سرطان کولون برحسب مرحله پاتولوژیک تومور

مرحله پاتولوژیک تومور	میزان مرگ ومیر ناشی از سرطان کولون	
	فراوانی مرگ و میر / فراوانی کل	میزان خطر (٪)
I	۱۶/۸۴	٪۱۹/۰
II	۷۳/۳۴۷	٪۲۱/۰
III	۹۴/۳۳۳	٪۲۸/۲
IV	۵۵/۹۸	٪۵۶/۱

برای نمایش دو متغیری متغیرهای رسته‌ای نیز می‌توان از جداول فراوانی و درصدی و نمودارهای میله‌ای در دو قالب خوشه‌ای (Cluster) و پشته‌ای (Stack) استفاده کرد. به عنوان مثال، درصد مرگ و میر ناشی از سرطان‌های کولون و رکتال بر حسب درجه‌های (Grades) مختلف تومور در نمودار ۳ نشان داده شده است.



نمودار ۳- نمایش فراوانی و درصد افراد مبتلا به سرطان‌های کولون و رکتال در رده‌های درجه تومور با استفاده از نمودارهای درصدی

همان‌طور که ملاحظه می‌شود، در نمودار میله‌ای خوشه‌ای قابلیت مقایسه شاخص‌های کنار هم در هر خوشه (سرطان‌های کولون و رکتال و همچنین درجه های مختلف تومور) وجود دارد حال آن که در نمودار پشته‌ای، علاوه بر تفکیک درصد مشاهدات به ازای متغیر اول (سرطان) امکان مشاهده‌ی کل مقدار درصدها به ازای متغیر درجه تومور (متغیر دوم) نیز وجود خواهد داشت.

در تعمیم حالت دو متغیره به بیش از دو متغیره، کافیست تفکیک دیگری به ازای رده‌های متغیرهای بعدی انجام شود. به عنوان مثال، مرگ ناشی از سرطان های کولون و رکتال (متغیر اول) برای رده‌های متغیر سیگار (متغیر دوم) به تفکیک رده‌های متغیر مصرف الکل (متغیر سوم) در جدول ۵ نشان داده شده است [۸].

[Downloaded from journals.tums.ac.ir on 2024-11-16]

جدول ۵- فراوانی و درصد وضعیت مرگ ومیر افراد مبتلا به سرطان کولورکتال برحسب رده‌های متغیرهای وضعیت کشیدن سیگار

وضعیت مصرف الکل	وضعیت کشیدن سیگار	وضعیت مرگ ومیر			
		زنده		مرده	
		فراوانی	درصد	فراوانی	درصد
هرگز مصرف نکرده	هرگز مصرف نکرده	۵۸۴	٪۷۳/۴	۲۱۲	٪۲۶/۶
	در حال حاضر یا قبلا مصرف کرده	۱۳۹	٪۶۶/۲	۷۱	٪۳۳/۸
در حال حاضر یا قبلا مصرف کرده	هرگز مصرف نکرده	۲۱	٪۷۲/۴	۸	٪۲۷/۶
	در حال حاضر یا قبلا مصرف کرده	۴۵	٪۶۵/۲	۲۴	٪۳۴/۸

خلاصه‌سازی روابط بین متغیرها در حالت چند متغیره (معیارهای توصیفی ارزیابی رابطه)

از این معیارها به صورت توصیفی برای خلاصه‌سازی رابطه میان متغیرها استفاده می‌شود. با توجه به این که ارائه‌ی این بحث نیاز به آشنایی با مفاهیم دیگری دارد، مطالب مربوط به آن در بخش های آتی این سری ارائه خواهد شد.

خلاصه‌سازی و نمایش داده‌ها برای متغیرهای کمی (عددی)

شیوه خلاصه‌سازی و گزارش نمودن متغیرهای عددی شامل مجموعه متغیرهای با مقیاس فاصله‌ای و نسبتی بر اساس جداول و نمودارهای مبتنی بر شاخص‌های آماری خاص خود صورت می‌گیرد که در این بحث بدان پرداخته می‌شود.

خلاصه‌سازی داده‌ها برای متغیرهای کمی

برای خلاصه نمودن ویژگی‌های مختلف داده‌های عددی از شاخص‌های مختلفی استفاده می‌شود. شاخص‌ها اعداد یا کدهایی هستند که ویژگی معینی در داده‌ها را تعریف می‌کنند. شاخص‌های خلاصه‌سازی داده‌ها شامل شاخص‌های مرکزی، پراکندگی، مکان و توزیع هستند که ۴ ویژگی مختلف در داده‌ها را خلاصه می‌نمایند.

شاخص‌های مرکزی

این شاخص‌ها معرف مرکز داده‌ها هستند و شامل شاخص‌های میانگین، میانه، و مد می‌باشند. میانگین مرکز ثقل داده‌ها، میانه نقطه وسط داده‌ها و مد (نما) عدد با بیشترین فراوانی است که به صورت زیر محاسبه می‌شوند:

میانگین = (مجموع مشاهدات) ÷ (تعداد مشاهدات)

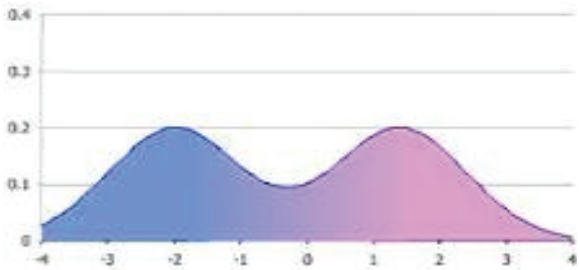
میانه: برای محاسبه میانه داده‌ها به ترتیب از کوچک به بزرگ مرتب شده عدد وسط در این داده‌های مرتب شده میانه است و در صورتی که تعداد مشاهدات زوج باشد، دو عدد در وسط قرار می گیرند، که در این صورت میانه برابر میانگین این دو عدد خواهد شد.

مد: برای محاسبه مد فراوانی هر یک از مشاهدات مشخص می‌شود؛ مد عددی (اعدادی) است که بیشترین فراوانی را داشته باشد. لازم به ذکر است در یک مجموعه از داده‌ها ممکن است بیشتر از یک مد وجود داشته باشد.

برای داده‌های کمی به عنوان اولین پیشنهاد، شاخص مرکزی میانگین توصیه می‌شود مگر این که در داده‌ها مشاهدات پرت وجود داشته باشد. مشاهدات پرت به داده‌هایی گفته می‌شود که از مجموعه داده‌ها فاصله دارند یا از آنها خیلی بزرگتر یا خیلی کوچکتر هستند. ملاک تشخیص پرت بودن داده‌ها، بر اساس شاخص‌های توزیع در بخش ۴.۱.۴ ارائه شده است. چون شاخص میانگین به حضور این گونه از مشاهدات حساس است، بنابراین در این موارد شاخص میانه پیشنهاد می‌شود که نسبت به داده‌های پرت اصطلاحاً نیرومند (Robust) است. البته میانه را غیر از چنین موقعیت‌هایی، برای خلاصه‌سازی متغیرهای رتبه‌ای با طیف وسیع نظیر اندازه‌گیری خستگی یا درد با استفاده از VAS نیز می‌توان به کار برد (Rezaei و همکاران [۶]، (a)). شاخص مد را زمانی که هدف یافتن مشاهده با بیشترین فراوانی باشد حتی برای متغیرهای با

[Downloaded from journals.tums.ac.ir on 2024-11-16]

مقیاس اسمی و رتبه‌ای نیز می‌توان به کار برد. برای داده‌های کمی، عمدتاً این شاخص در مجموعه‌هایی که چند نقطه اوج فراوانی در آنها وجود دارد (چند مدی) گزارش می‌شود (شکل ۱).



شکل ۱- توزیع فراوانی دو مدی

شاخص‌های پراکندگی

این شاخص‌ها معرف پراکندگی موجود در داده‌ها می‌باشند و شامل شاخص‌های دامنه تغییرات، انحراف معیار، ضریب تغییرات و دامنه‌ی میان‌چارکی هستند. دامنه‌ی تغییرات حد فاصل بین کوچکترین و بزرگترین مشاهده می‌باشد. انحراف معیار جذر واریانس است که واریانس برابر متوسط فاصله داده‌ها از شاخص میانگین می‌باشد. ضریب تغییرات، پراکندگی نسبی داده‌ها نسبت به شاخص میانگین و دامنه میان چارکی حد فاصل بین صدک های ۲۵ و ۷۵ است که در آن صدک‌های ۲۵ و ۷۵ به ترتیب اعدادی هستند که ۲۵ و ۷۵ درصد داده‌ها از آن کوچک‌ترند. بحث کامل در مورد این شاخص در بخش شاخص‌های ممکن ارائه خواهد شد. شاخص‌های فوق به صورت زیر محاسبه می‌شوند:

دامنه تغییرات = بزرگترین مشاهده – کوچکترین مشاهده

واریانس = (مشاهده اول منهای میانگین)^2 + ... + (مشاهده آخر منهای میانگین)^2) / (تعداد مشاهدات منهای یک)

انحراف معیار = جذر واریانس

شایان ذکر است که باید در کنار هر شاخص مرکزی یک شاخص پراکندگی گزارش شود تا میزانی از پراکندگی در داده‌ها در اطراف آن شاخص مرکزی مشخص شده باشد. به عنوان مثال، در اندازه‌گیری وزن طلا اگر از ترازویی استفاده شود که میزان دقت آن برحسب ۰/۱ گرم باشد در این صورت ممکن است یک قطعه طلای ۵ گرمی را در دامنه ۴/۹ تا ۵/۱ گرم گزارش نماید (۰/۱ gr ± ۵)، اما اگر میزان دقت میزان دقت آن برحسب ۰/۰۱ گرم باشد در این صورت ممکن است یک قطعه طلای ۵ گرمی را در دامنه ۴/۹۹ تا ۵/۰۱ گرم گزارش نماید (۰/۰۱ gr ± ۵) و مسلم است که میزان اعتماد به ترازوی دوم بیشتر است. اما همان‌طور ملاحظه می‌شود، در گزارش هر دو ترازو، عدد ۵ گرم به عنوان نقطه وسط گزارش خواهد شد. به عبارت دیگر میزان اعتماد به ۵ گرمی‌که در اطراف آن تغییرات ۰/۰۱ وجود دارد در مقایسه با ۵ گرمی‌که در اطراف آن تغییرات ۰/۱ وجود دارد، بیشتر است. خلاصه این که نمی‌توان شاخص مرکزی را به تنهایی برای گزارش یک سری از داده‌ها استفاده نمود و نیاز است در کنار آن یک شاخص پراکندگی میزانی از دقت یا اعتمادپذیری شاخص مرکزی مورد نظر را فراهم نماید.

دامنه‌ی تغییرات شاخص ساده‌ای به لحاظ محاسباتی است ولی به دلیل این که از همه داده‌ها استفاده نمی‌کند، به مشاهدات انتهایی در داده‌ها بستگی دارد و به سادگی دچار نوسان می‌شود، معمولاً مورد نقد است ولی می‌توان آن را به عنوان اندازه‌ای از تفاوت بین حداقل و حداکثر مشاهدات به همراه میانگین و انحراف معیار و یا در کنار میانه گزارش نمود.

واریانس، در واقع برآوردی از متوسط فاصله آماری داده‌ها نسبت به مرکز ثقل آنها فراهم می‌کند، اما با توجه به این که برای محاسبه این شاخص از توان ۲ استفاده می‌شود، بعد اندازه‌گیری این شاخص نیز توان ۲ خواهد داشت (مثلاً (وزن)^2). در این صورت از این لحاظ با میانگین (که بعد اندازه‌گیری آن مشابه مشاهدات است) تجانس ندارد. جذر گرفتن از واریانس که شاخص انحراف معیار را نتیجه

[Downloaded from journals.tums.ac.ir on 2024-11-16]

می‌دهد، این مشکل را حل می‌کند. به همین دلیل اگر برای مجموعه‌ای از داده‌ها شاخص مرکزی میانگین گزارش شود، در کنار آن شاخص پراکندگی انحراف معیار نیز گزارش می‌شود.

ضریب تغییرات، پراکندگی نسبی اندازه‌گیری یک متغیر نسبت به میانگین خودش را نشان می‌دهد (انحراف معیار تقسیم بر میانگین) و با توجه به این که کمیتی بدون بعد اندازه‌گیری است، برای مقایسه دو اندازه‌گیری با ابعاد مختلف نظیر فشار خون و ضربان قلب، قابل به کارگیری است. همچنین این شاخص برای مقایسه‌ی اندازه‌گیری‌های یکسان در دو جمعیت متفاوت نیز به کار می‌رود.

شاخص‌های مکان

این شاخص‌ها معرف مکان داده‌ها و شامل میانه، سه یک‌ها- چارک‌ها- پنج یک‌ها (پنجک‌ها)- ...، می‌باشند؛ میانه عددی است که در مکان وسط مشاهدات قرار می‌گیرد و چون شاخصی است مکان وسط در داده‌ها را نشان می دهد، به عنوان یک شاخص مرکزی نیز از آن استفاده می‌شود. سه یک اول و دوم به ترتیب اعدادی هستند که یک سوم و دو سوم داده‌ها از آن کوچک‌ترند. به همین ترتیب چارک اول، دوم و سوم به ترتیب اعدادی هستند که یک چهارم (۲۵٪)، دو چهارم (نصف یا ۵۰٪) و سه چهارم (۷۵٪) مشاهدات از آنها کوچکترند و .... برای محاسبه شاخص‌های مکان، ابتدا داده‌ها به ترتیب از کوچک به بزرگ مرتب می‌شوند، سپس مکان شاخص مورد نظر بر اساس درصدی که در تعریف آن شاخص وجود دارد مشخص شده و در نهایت مقدار شاخص مشخص می‌شود. مثلاً برای یافتن صدک ۲۵، پس از مرتب کردن داده‌ها به طور صعودی، مکان متناظر با این شاخص در نقطه درصدی ۲۵ مشخص شده و عدد موجود در این مکان، همان صدک ۲۵ می‌باشد. البته اگر مکان مزبور بین دو مشاهده قرار گیرد از آن دو مشاهده میانگین وزنی گرفته می‌شود و به عنوان شاخص مکان مورد نظر ارائه می‌شود.

یکی از کاربردهای شاخص‌های مکان، برای تعیین وضعیت و سطوح اقتصادی اجتماعی و یا مشخص کردن دهک‌های درآمدی در جامعه در مطالعات اقتصادی می‌باشد. دهک اول در آمدی در جامعه عددی است که در آمد ۱۰ درصد از افراد جامعه زیر آن عدد قرار می‌گیرند.

در بررسی روند رشد کودکان نیز کاربرد دیگری از این شاخص‌ها را می‌توان ملاحظه نمود؛ در این زمینه صدک های ۳ و ۹۷ بسیار پرکاربرد هستند صدک ۵۰ یا میانه نقطه وسط روند رشد کودکان (از لحاظ وزن، قد و دور سر) را نشان می‌دهد که نقطه‌ی ایده آل به ازای سنین مختلف رشدی است. صدک‌های ۳ و ۹۷ برای تعیین وضعیت‌های بحرانی روند رشد کاربرد دارند؛ به طوری که کودکان با رشد زیر صدک ۳ و بالای صدک ۹۷ وضعیت بحرانی دارند. با این دیدگاه، در حالت کلی نیز برای یافتن وضعیت‌های غیر عادی از شاخص‌های مکان می‌توان استفاده کرد.

یکی دیگر از کاربردهای شاخص‌های مکان، برای دسته‌بندی مشاهدات کمی است؛ زمانی که ملاک زمینه‌ای یا بالینی برای دسته‌بندی متغیرهای کمی وجود داشته باشد، نظیر دسته‌بندی «۱۸/۵ >» = «کم وزن»، «۲۵-۱۸/۵ <» = «نرمال»، «۳۰-۲۵» = «اضافه وزن»، و « ۳۰ <» = «چاق» برای نمایه توده بدنی در افراد بزرگسال، پیشنهاد می‌شود که از این دسته‌بندی‌های زمینه‌ای استفاده شود. ولی زمانی که ملاک زمینه‌ای وجود نداشته باشد، شاخص‌های مکان نقطه برش های مطلوبی برای دسته‌بندی مشاهدات فراهم می‌کنند که بر اساس دسته‌بندی‌های حاصل از این شاخص‌ها معمولاً تعداد متعادلی از مشاهدات در این زیر دسته‌ها ایجاد می‌شود که تحلیل‌های بعدی روی این دسته‌ها را دچار نوسان نمی‌کند. دسته‌بندی مشاهدات در ۲، ۳، ۴ یا ۵ دسته به ترتیب با استفاده از شاخص‌های میانه، سه یک، چارک و پنج یک، نقطه برش مورد نظر را می‌توان به دست آورد. دامنه میان چارکی به صورت زیر محاسبه می‌شود:

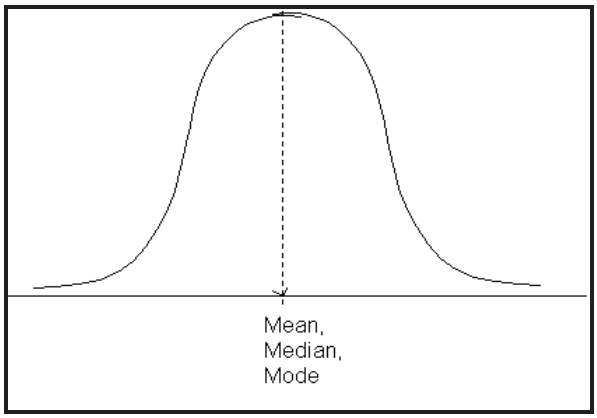
دامنه میان چارکی = صدک ۷۵ – صدک ۲۵

و در واقع دامنه‌ای را شامل می‌شود که ۵۰٪ میانی مشاهدات در این دامنه قرار می‌گیرند. از این شاخص می‌توان در کنار میانه (علاوه بر دامنه تغییرات) به عنوان یک شاخص پراکندگی استفاده کرد.

[Downloaded from journals.tums.ac.ir on 2024-11-16]

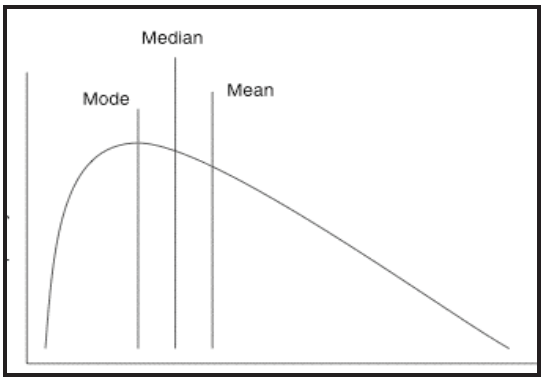
شاخص‌های توزیع

این شاخص‌ها معرف شکل توزیعی در داده‌ها و شامل شاخص‌های کجی (Skewness) و کشیدگی (Kurtosis) هستند. برای روشن شدن این مفاهیم، یک امتحان درسی استاندارد را در نظر بگیرید؛ در این امتحان انتظار می‌رود که اغلب نمرات در اطراف یک شاخص مرکزی نظری میانگین (مثلاً ۱۶) قرار گیرد و تجمع نمرات در اطراف میانگین زیاد باشد و هرچه از میانگین فاصله ایجاد می‌شود (یعنی به سمت ۲۰ و ۱۲) تعداد افرادی که این نمرات را کسب می‌کنند کمتر و کمتر می‌شود (شکل۲).



شکل ۲- نمایش توزیع متقارن نرمال

این نمودار اصطلاحاً توزیع نرمال (طبیعی) نامیده می‌شود که شکلی متقارن و شبیه زنگ دارد (زنگدیس یا Bell-Shaped). در این توزیع میانگین، میانه و مد روی یک نقطه قرار می‌گیرند و در این توزیع مقادیر شاخص‌های کجی و کشیدگی برابر صفر است. حال یک امتحان سخت را در نظر بگیرید، در این صورت انتظار می‌رود که اغلب افراد نمرات پایین‌تر داشته باشند و تعداد افرادی که نمرات بالا می‌گیرند، کم است و منحنی به صورت زیر ایجاد می‌شود که نامتقارن است (شکل ۳).

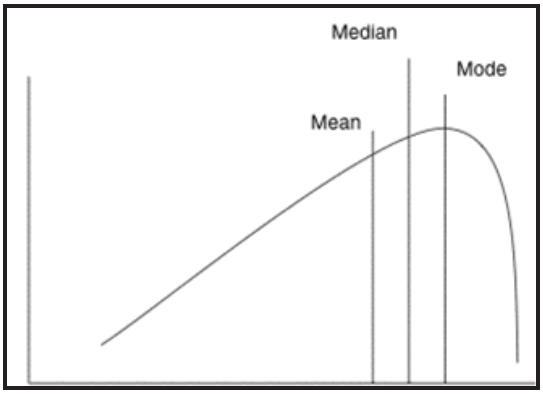


شکل ۳- نمایش توزیع کج به راست (چوله به راست)

در این شکل که دامنه یا دنباله توزیع به سمت راست امتداد یافته تا داده‌های بزرگ را پوشش دهد، اصطلاحاً چوله به راست است یا کجی مثبت دارد که در آن ترتیب میانگین> میانه > مد را می‌توان مشاهده نمود. در این حالت مقدار کجی مثبت خواهد شد.

برعکس حالت در یک امتحان راحت انتظار می‌رود که اغلب افراد نمرات بالاتر داشته باشند و تعداد افرادی که نمرات پایین می‌گیرند، کم است و منحنی به صورت زیر ایجاد می‌شود که نامتقارن است (شکل ۴):

[Downloaded from journals.tums.ac.ir on 2024-11-16]

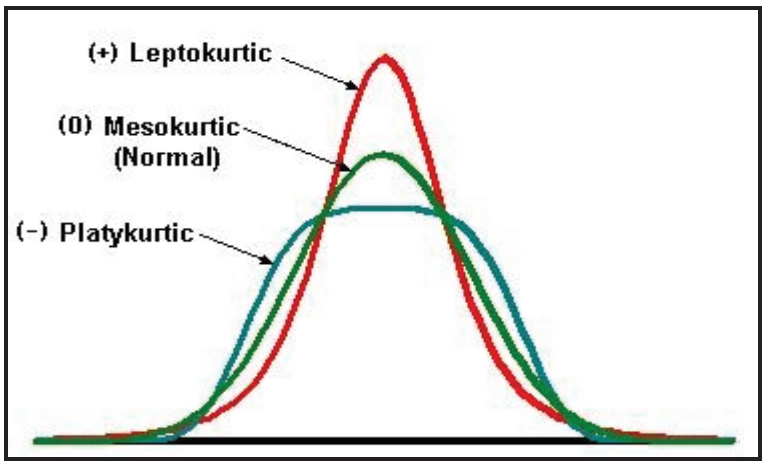


شکل ۴- نمایش توزیع کج به چپ (چوله به چپ)

در این شکل که دامنه یا دنباله توزیع به سمت چپ امتداد یافته تا داده‌های کوچک را پوشش دهد، اصطلاحاً چوله به چپ است یا کجی منفی دارد که در آن ترتیب میانگین< میانه < مد را می‌توان مشاهده نمود. در این حالت مقدار کجی منفی خواهد شد.

کج بودن توزیع به عنوان ملاکی برای انحراف از تقارن (نرمال)، سبب می‌شود که میانگین (انحراف معیار) شاخص مناسبی برای گزارش کردن داده‌ها نباشد. به عبارت دیگر کج بودن توزیع، حضور داده‌های پرت و پراکندگی نامتقارن موجود در داده‌ها را نشان می‌دهد که در چنین مواردی میانه نسبت به میانگین شاخص بهتری است. در حالت ایده‌ال توزیع متقارن (نرمال) کجی صفر دارد، ولی در عمل مقادیر در دامنه (۱/۵ – تا ۱/۵) مورد قبول است و دامنه (۳- تا ۳) به عنوان حد فرین در نظر گرفته می‌شود یعنی کجی خارج از این دامنه انحراف شدید از نرمال را نشان می‌دهد. به عبارت دیگر، در صورت وجود مشاهدات پرت توزیع داده‌ها غیرنرمال خواهد شد که در این صورت برای خلاصه سازی متغیرهای کمی غیر نرمال، به جای میانگین (انحراف معیار) باید از میانه (صدک ۲۵ – صدک ۷۵) استفاده شود.

شاخص کشیدگی، ارتفاع قله توزیع نسبت به یک ملاک را نشان می‌دهد. به عبارت دیگر، هر چه تمرکز مشاهدات در اطراف میانگین بیشتر (کمتر) باشد در این صورت ارتفاع منحنی در این نقطه از یک وضعیت طبیعی بیشتر (کمتر) خواهد بود که در این صورت مقدار شاخص کشیدگی به ترتیب بزرگتر (کوچکتر) از صفر خواهد شد (شکل ۵). در عمل مقادیر در دامنه (-۲ تا ۲) مورد قبول است و دامنه (۱۰- تا ۱۰) به عنوان حد فرین در نظر گرفته می‌شود یعنی کشیدگی خارج از این دامنه انحراف شدید از نرمال را نشان می‌دهد.



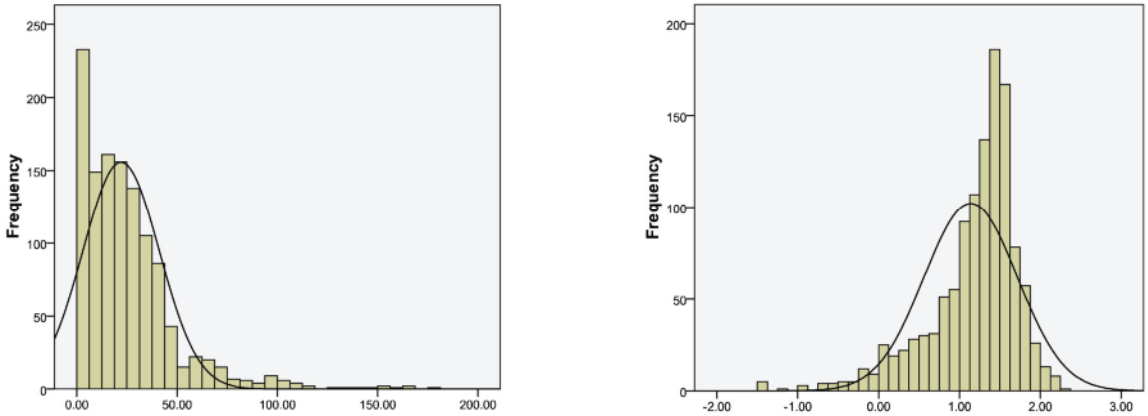
شکل ۵- نمایش منحنی‌های با کشیدگی‌های مختلف

[Downloaded from journals.tums.ac.ir on 2024-11-16]

نمایش داده‌ها در حالت یک متغیره

یافته‌های حاصل برای یک متغیر کمی را می‌توان در قالب جدول یا نمودار ارائه کرد. برای نمایش یافته‌ها در جدول معمولاً به صورت کلی و بدون تفکیک و غالبا برای متغیرهای دموگرافیک کمی به صورت میانگین (انحراف معیار) یا میانه (صدک ۲۵ – صدک ۷۵) صورت می‌گیرد. نمودارهایی که برای نمایش متغیر کمی به صورت انفرادی به کار می‌روند، شامل:

۱) هیستوگرام (بافت‌نگار یا مستطیلی) برای نمایش مرکز، پراکندگی و توزیع. به عنوان مثال، برای متغیر زمان پیگیری افراد در داده‌های مربوط به سرطان کولورکتال نمودار ۴ حاصل شده است. برای داده‌های اصلی این منحنی دارای مدی تقریباً برابر ۲۵ و پراکندگی بیشتر در طرف راست منحنی می‌باشد و در نتیجه دارای توزیع کج به راست است. در چنین داده‌هایی استفاده از تبدیلات نظیر لگاریتم گرفتن از داده‌ها می‌تواند تا حدودی شکل داده‌ها را به سمت متقارن متمایل نماید.



نمودار ۴- نمایش متغیر زمان پیگیری افراد (سمت چپ) و لگاریتم آن (سمت راست) توسط هیستوگرام (با منحنی نرمال)

۲) نمودار چندضلعی (چندبر یا Polygon) برای نمایش مرکز، پراکندگی، توزیع داده‌ها (و مقایسه این شاخص‌ها در دو یا چند گروه برای حالت دو متغیره). به عنوان مثال، برای متغیر زمان پیگیری افراد این نمودار را مشابه با هیستوگرام می توان رسم نمود. این نمودار را می‌توان به تفکیک گروه‌ها مثلاً به تفکیک دو گروه زنان و مردان نیز رسم نمود.

۳) نمودار ساقه و برگ برای نمایش مرکز، پراکندگی، توزیع داده‌ها و در عین حال بر حسب مقادیر داده‌ها. به عنوان مثال، برای متغیر زمان پیگیری افراد در داده‌های مربوط به سرطان کولورکتال نمودار ۶ حاصل شده است. مشابه هیستوگرام، برای داده‌های اصلی این منحنی دارای مدی تقریباً برابر ۲۵ و پراکندگی بیشتر در طرف راست منحنی می باشد و در نتیجه دارای توزیع کج به راست است. به علاوه مقادیر داده‌ها نیز در این نمودار مشخص هستند که به اظهار نظر دقیق‌تر در مورد مرکز، پراکندگی وشکل توزیعی داده‌ها کمک می‌کند.

[Downloaded from journals.tums.ac.ir on 2024-11-16]

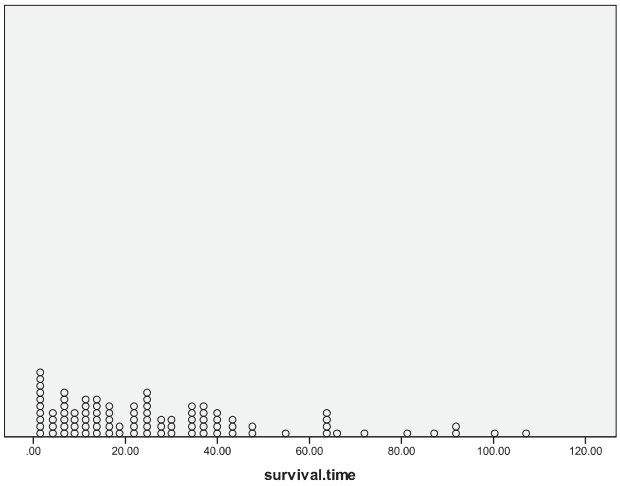
survival.time Stem-and-Leaf Plot

Frequency	Stem & Leaf
25.00	0 . 0000112222345556667778889
19.00	1 . 0001122334445666788
16.00	2 . 0122234444456789
13.00	3 . 0033445667779
8.00	4 . 00022468
1.00	5 . 4
5.00	6 . 23445
1.00	7 . 1
1.00	8 . 1
5.00 Extremes	(>=87)

Stem width: 10.00
Each leaf: 1 case(s)

نمودار ۵- نمایش متغیر زمان پیگیری افراد توسط نمودار ساقه و برگ

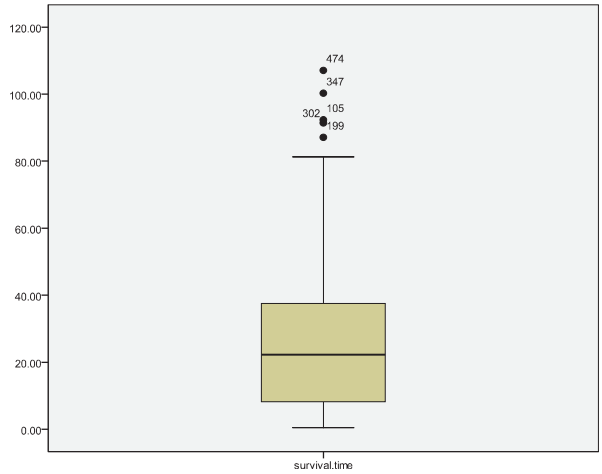
۴) نمودار نقطه‌ای برای نمایش مرکز، پراکندگی و تشخیص نقاط پرت احتمالی داده‌ها. به عنوان مثال، برای متغیر زمان پیگیری افراد در داده‌های مربوط به سرطان کولورکتال نمودار ۷ حاصل شده است. در این نمودار نقاط پرت موجود در این مطالعه در انتهای محور افقی قابل مشاهده هستند.



نمودار ۶- نمایش متغیر زمان پیگیری افراد توسط نمودار نقطه‌ای

۵) نمودار جعبه‌ای برای نمایش مرکز، مکان، پراکندگی، توزیع داده‌ها و برای زمانی که در داده‌ها مشاهدات پرت وجود دارد و نمایش این داده‌ها توسط میانگین مناسب نمی‌باشد. به علاوه این نمودار را برای نمایش داده‌های رتبه‌ای با طیف وسیع نیز می‌توان به کار برد. به عنوان مثال، زمان بقای افراد در مطالعه توسط نمودار نشان داده شده است. در این نمودار میانه، چارک اول، چارک دوم، مینیمم و ماکسیمم (خلاصه ۵ عددی) قابل مشاهده است. علاوه بر آن نقاط پرت بزرگ در این نمودار در انتهای بالایی نمودار مشاهده می شود. لازم به ذکر است که در این نمودار مقادیر مینیمم و ماکسیمم پس از کنار گذاشتن این نقاط پرت محاسبه می‌شوند.

[Downloaded from journals.tums.ac.ir on 2024-11-16]



نمودار ۷- نمایش متغیر زمان پیگیری افراد توسط نمودار جعبه‌ای

خلاصه‌سازی و نمایش داده‌ها در حالت چند متغیره

منظور از حالت چند متغیره، خلاصه‌سازی و نمایش هم‌زمان چند متغیر است که معمولاً به دلیل بررسی رابطه دو یا چند متغیر و بررسی اختلاف یا یا چند متغیر در رده‌های متغیر(های) دیگر مورد بررسی قرار می‌گیرد. در این حالت، ممکن است هدف بررسی رابطه‌ی یک یا چند متغیر کمی با یک یا چند متغیر کمی، رتبه‌ای یا اسمی باشد که با تاکید بر توصیف و نمایش این روابط (و نه آزمون کردن آنها)، در این بخش مورد بررسی قرار خواهند گرفت و برای سادگی به صورت دو متغیره بررسی خواهند شد و تعمیم آنها برای حالت بیش از دو متغیر سر راست خواهد بود.

خلاصه‌سازی داده‌ها در حالت دو متغیره

خلاصه‌سازی یک متغیر کمی به ازای رده‌های یک متغیر کیفی (وضعیت کمی- کیفی (اسمی یا ترتیبی)

در چنین وضعیتی مشابه با حالت یک متغیره، یک از شاخص‌های میانگین (انحراف معیار) یا میانه (صدک ۲۵ – صدک ۷۵) برای متغیر کمی مورد بررسی محاسبه می‌شود. با این تفاوت که شاخص‌های فوق، به تفکیک رده‌های متغیر کیفی (اسمی- ترتیبی) مورد بررسی محاسبه می‌شوند. به عنوان مثال، شاخص‌های خلاصه برخی از متغیرهای مطالعه بررسی اثر پروبیتیک‌ها روی شاخص‌های لیپیدی در بیماران مبتلا به دیابت نوع ۲ در جدول زیر آورده شده است.

جدول ۶- شاخص‌های خلاصه متغیرهای مطالعه به تفکیک دو گروه دریافت کننده ماست معمولی و پروبیوتیک

متغیر	گروه دریافت کننده ماست معمولی	گروه دریافت کننده ماست پروبیوتیک
سن (سال) #	۵۱/۰۰ (۷/۳۴)	۵۰/۸۷ (۷/۶۷)
وزن #	۷۵/۴۲ (۱۱/۲۸)	۷۶/۱۸ (۱۰/۹۶)
متفورمین (تعداد قرص) $	۲/۰۰ (۱/۲۵)	۲/۰۰ (۱/۲۵)
گلی بنکلامید(تعداد قرص) $	۱/۰۰ (۱/۰۰)	۲/۰۰ (۲/۰۰)

#: میانگین (انحراف معیار)

$: میانه (دامنه میان چارکی)

منبع: [۹]

[Downloaded from journals.tums.ac.ir on 2024-11-16]

خلاصه‌سازی یک متغیر کمی به ازای رده‌های یک متغیر رتبه‌ای

در چنین وضعیتی، علاوه بر این که می‌توان مشابه وضعیت قبلی عمل کرد، برای توصیف ارتباط بین این دو متغیر می‌توان شاخص‌همبستگی رتبه‌ای اسپیرمن یا تاو (τ) کندال را محاسبه نمود. شاخص‌های فوق به عنوان توصیفی از رابطه بین دو متغیر رتبه‌ای نیز به کار می‌روند و مقادیر این شاخص بین ۱- تا ۱+ قرار می گیرد که در آن ۱- رابطه کامل معکوس و ۱+ رابطه کامل مستقیم و صفر عدم رابطه بین دو متغیر را نشان می‌دهد. توضیح مفصل بررسی رابطه‌ی بین متغیرها در بخش آزمون‌های بررسی رابطه بین متغیرها ارائه خواهد شد.

خلاصه‌سازی یک متغیر کمی به همراه یک متغیر کمی دیگر

برای توصیف ارتباط بین این دو متغیر می‌توان شاخص همبستگی پیرسون را محاسبه نمود. تفسیر این شاخص مشابه شاخص همبستگی رتبه‌ای اسپیرمن یا تاو (τ) کندال می‌باشد.

نمایش داده‌ها در حالت دو (چند) متغیره

این بحث نیز در حالت دو متغیره توضیح داده خواهد شد ولی تعمیم آن به بیش از دو متغیر سر راست می‌باشد. برای گزارش یک متغیر کمی به ازای رده‌های یک متغیر کیفی (وضعیت کمی- کیفی (اسمی یا ترتیبی) می‌توان شاخص‌های مناسب را در قالب جدول یا نمودار ارائه کرد. در جداول میانگین (انحراف معیار)، میانه (صدک ۲۵ – صدک ۷۵) به تفکیک رده‌های متغیر کیفی ارائه می‌شود. البته در صورتی که نیاز باشد، دامنه مشاهدات نیز ارائه شود، در این صورت مقادیر مینیمم و ماکسیمم داده‌ها نیز ارائه خواهد شد. برای ارائه شاخص‌های فوق باید تعداد مناسب از ارقام اعشار (معمولا دو رقم) برای هر دو شاخص مرکزی و پراکندگی انتخاب شود.

مأخذ

1- Asghari-Jafarabadi M, Hajizadeh E, Kazemnejad A, Fatemi SR. Site-Specific Evaluation of Prognostic Factors in Iranian Colorectal Cancer Patients: A Competing Risks Survival Analysis. *Asian Pacific Journal of Cancer Prevention* 2009; 10:815-22.

2- Allahverdipour H, Asghari-Jafarabadi M, Emami A. Breast cancer risk perception, benefits of and barriers to mammography adherence among a group of Iranian women. *Women Health* 2011 51(3):204-19.

3- Agresti A. Categorical data analysis. New York: John Wiley & Sons 2002. PP 3-4.

4- Asghari-Jafarabadi M, Hajizadeh E, Kazemnejad A, Fatemi SR. Comparison the Role of BMI, Pathologic Stage and Hereditary Related Factors on Survival between Colon and Rectal Cancers: Frailty Competing Risks Model. *Iranian Journal of Epidemiology* 2010;6(3):36-50. [in Persian]

5- Johnson DR, Creech JC. Ordinal measures in multiple indicator models: A simulation study of categorization error. *American Sociological Review* 1983; 48:398-407.

6- Rezaei Adaryani M, Ahmadi F, Mohamadi E, Asghari Jafarabadi M. The effect of three positioning methods on patient outcomes after cardiac catheterization. *J of Advanced Nursing* 2009; 65(2):417-24.

7- Asghari-Jafarabadi M, Mohammadi SM, Hajizadeh E, Kazemnejad A, Fatemi SR. Does the Prognosis of Colorectal Cancer Vary with Tumor Site? *Gastroenterology and Hepatology from Bed to Bench* 2011; 4(4):199-209.

8- Asghari-Jafarabadi M, Hajizadeh E, Kazemnejad A, Fatemi SR. A comparative study on the prognostic impact of concurrent smoking and alcohol drinking on colon and rectal cancers: A frailty competing risks survival analysis. *Gastroenterology and Hepatology from Bed to Bench* 2010; 3(1):19-26.

9- Ejtahed HS, Mohtadi-Nia J, Homayouni-Rad A, Niafar M, Asghari Jafarabadi M, Mofid V, et al. Effect of Probiotic Yoghurt Containing Lactobacillus acidophilus and Bifidobacterium lactis on Lipid Profile in Individuals with Type 2 Diabetes Mellitus. *Journal of Dairy Science* 2011 94(7):3288-94.

[Downloaded from journals.tums.ac.ir on 2024-11-16]