

نکاتی در خصوص کاربرد نمودارها در گزارش‌های آماری

رحیم محمودوند
گروه آمار، دانشگاه بوعلی سینا همدان

چکیده

نمودارها از مهم‌ترین و ساده‌ترین ابزارهای آمار توصیفی هستند که معمولاً در تهیه‌ی گزارش‌ها و انجام تحلیل‌های آماری مورد استفاده قرار می‌گیرند. تنوع داده‌ها و استفاده از آن‌ها باعث به وجود آمدن نمودارهای مختلفی شده است که برخی از آن‌ها بسیار ساده و قابل فهم و برخی دیگر از لحاظ ترسیم و فهم برای کاربران پیچیده‌تر و دشوارتر هستند. در این میان چنانچه کاربران با ویژگی‌های این نمودارها و نحوه‌ی تفسیر آن‌ها به اندازه کافی آشنا نباشند ممکن است در انتخاب صحیح نمودارها دچار سردرگمی شده و از کیفیت گزارش بکاهند. در این مقاله به برخی از مواردی که معمولاً در استفاده از نمودارها و به ویژه در تدریس آن‌ها کمتر مورد توجه هستند، اشاره شده است. انتخاب نوع نمودار، انتخاب شیوه‌ی نمایش نمودار و تفسیر نتایج از جمله مواردی است که در این مقاله بیشتر مورد بحث قرار گرفته است. علاوه بر این مرور تاریخچه مختصری در خصوص روند معرفی نمودارهای آماری و نقدهایی که در کاربرد آن‌ها مطرح شده است نیز در مقاله ارائه شده است.

واژه‌های کلیدی: آمار توصیفی، نمودارهای آماری، گزارش‌های آماری.

۱ مقدمه

پژوهش را در قالبی قابل فهم خلاصه نمائیم. به این منظور معمولاً در مباحث آمار توصیفی از سه حوزه‌ی زیر استفاده می‌کنند:

- رسم نمودارهای آماری،
- تهیه‌ی جداول فراوانی،
- محاسبه‌ی شاخص‌های آماری.

زمانی که یک پژوهش علمی به مرحله عملی می‌رسد، مجموعه‌ای از اطلاعات، که یا به شکل عدد و رقم هستند یا به عدد و رقم تبدیل می‌شوند، به دست می‌آید که ممکن است خیلی گیج‌کننده و غیر قابل فهم باشند. مطمئناً با دیدن مجموعه‌ای از اعداد و ارقام نمی‌توان به نتایج مفید

و قابل استفاده دست یافت. ولی از آنجایی که یکی از اهداف هر تحقیق علمی رسیدن به نتیجه‌ای کلی است، بنابراین نیاز داریم که داده‌های خام حاصل از تحقیق و به نحوی ترسیم شوند که بتوان به راحتی اطلاعات نهفته از لحاظ میزان سادگی فهم می‌توان گفت که ساده‌ترین ابزار همان نمودارهای آماری است که بنا به گفته محققین باید

متداول را با توجه به این نگاه نشان می‌دهد. بر اساس این جدول با توجه به نوع داده‌ها می‌توان نمودارهای آماری مناسبی پیشنهاد کرد. با این وجود ملاحظه می‌کنیم که

جدول ۱: دسته بندی نمودارهای آماری متداول

دایره‌ای میله‌ای	کیفی	یک متغیری	نمودارهای آماری
ساقه و برگ مستطیلی (بافت نگار) جعبه‌ای	کمی		
میله‌ای	کیفی	چند متغیری	
پراکنش پراکنش خطی	کمی ترکیبی		

برای هر حوزه چند نمودار به عنوان پیشنهاد وجود دارد. اما آیا استفاده از هر یک از آن‌ها در عمل مناسب است؟ همچنین در نحوه‌ی نمایش و تفسیر این نمودارها چه ملاحظه‌هایی را باید در نظر گرفت؟ متأسفانه در اغلب کتاب‌های آمار مقدماتی پاسخ مناسب و کاملی برای چنین سوال‌هایی مطرح نشده است و در بسیاری موارد به چنین موضوعی اصلاً ورود نکرده و تنها به معرفی این نمودارها و نحوه‌ی ترسیم آن‌ها پرداخته شده است. این در حالی است که امروزه با توجه به امکان ترسیم این نمودارها و بسیاری از انواع نمودارهای دیگر در نرم افزارهای کامپیوتری نیاز به آموزش نحوه‌ی ترسیم دستی نمودارها از اهمیت کمتری برخوردار است و در عوض مسائلی مانند آنچه در بالا گفته شد پراهمیت‌تر به نظر می‌رسند. کراسکال در سال ۱۹۷۵ بیان می‌دارد که به طور کلی نقش نمودارهای آماری در گسترش علوم آماری افت و خیزهای چشم‌گیری داشته است: در زمانی مانند دوره‌ی کارل پیرسن که از فرانسویس گالتون پیروی می‌کرده نمودارهای آماری هسته‌ی اصلی آمار بودند در حالیکه چند سال بعد از آن با شکوفایی

در داده‌ها را از روی آن تا حدودی با چشم بدون توضیح و تشریح اضافی دید (بهبودیان [۱]). البته برخی محققین نیز معتقدند که علاوه بر نحوه‌ی ترسیم نمودار از لحاظ فنی و به لحاظ قابلیت فهم اطلاعات درون آن، باید جنبه‌ی زیبایی و ظواهر نمودار را نیز در نظر داشت. این موضوع به ویژه از حدود سال ۱۹۷۰ به بعد با گسترش فناوری‌های کامپیوتری و امکان ترسیم نمودارها با کامپیوتر مورد توجه بیشتری قرار گرفته است. اما واضح است که برای فهم جدول فراوانی و شاخص‌های آماری، دانش بیشتری از آمار مورد نیاز است و لذا درک آن برای همگان به سادگی نمودار نیست.

البته با وجود سادگی نمودارها قابل ذکر است که عدم استفاده از آن‌ها و اتکاء به شاخص‌های آماری ممکن است در برخی شرایط نتیجه‌گیری‌های نادرستی ایجاد کند. یکی از معروف‌ترین مثال‌ها در این زمینه مربوط به مبحث ارتباط و همبستگی است که معمولاً بدون توجه به نتایج حاصل از نمودارهای آماری، از شاخص‌های همبستگی استفاده شده و غالباً نتایج نادرستی گزارش می‌شود. این در حالی است که با یک نمودار ساده می‌توان یک نتیجه‌ی دقیق‌تر ارائه کرد. در این خصوص به عنوان نمونه انسکامب [۳] یک مجموعه داده‌ی ساختگی در نظر گرفت که دارای میانگین و انحراف معیار برابری بودند و نشان داد که همه‌ی آن‌ها دارای ضریب همبستگی خطی برابر و نسبتاً بالایی هستند در حالیکه بین یک مجموعه از آن‌ها ارتباط ناخطی کاملاً واضحی وجود داشت.

نمودارها بر حسب اینکه داده‌های کیفی یا کمی را نمایش می‌دهند به دو دسته کیفی یا کمی تقسیم می‌شوند. بعلاوه بر حسب اینکه اطلاعات مربوط به یک، دو یا چند سری داده را نمایش می‌دهند به دسته‌های یک، دو یا چند متغیری تقسیم می‌شوند. جدول ۱ برخی از نمودارهای آماری

۲ نمودارهای آماری برای داده‌های کیفی

جنبه‌های ریاضی آمار و احتمال، نمودارهای آماری کم‌کم به حاشیه رانده شده و در بسیاری موارد نادیده گرفته شده است. اما اخیراً دوباره گرایش به سمت استفاده از نمودارهای آماری قوت گرفته و بسیاری از آماردانان به استفاده از نمودارهای آماری توصیه می‌کنند (بنیگر و رابین [۴]).

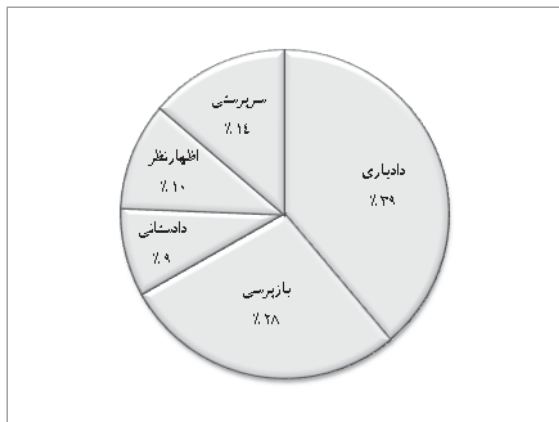
علاوه بر این کاکس [۵] نیز بیان می‌دارد که بحث نمودارهای آماری یک موضوع بین رشته‌ای است و در این خصوص باید ملاحظه‌های مختلفی را از منظر جنبه‌های علمی، آماری، محاسباتی، روانشناسی و جامعه‌شناسی در نظر گرفت. وی بیان می‌کند که کتاب‌های خوب درباره‌ی نمودارهای آماری بسیار کم است. به اعتقاد ایشان یک دلیل برای این موضوع آن است که تنها یک دانشمند همه‌چیزدان می‌تواند درباره‌ی یک موضوع چند وجهی مانند نمودارهای آماری قضاوت کند. بحث در خصوص مسائلی که در بالا به برخی از آن‌ها اشاره شد و نیز توصیه‌هایی در خصوص استفاده از نمودارهای رایج از اهداف اصلی این مقاله است.

مطالب این مقاله در ادامه به صورت زیر تنظیم شده است: در بخش بعدی دو نمودار کیفی متداول؛ یعنی نمودارهای دایره‌ای و میله‌ای با استفاده از چند مثال مورد مقایسه قرار گرفته و نکاتی که باید بدان در عمل توجه شود مورد بحث قرار می‌گیرند. در بخش سوم بحث مشابهی برای نمودارهای کمی که مشتمل بر بافت نگار (هیستوگرام)، نمودار ساقه و برگ و نمودار جعبه‌ای هست، مطرح می‌شود. در پایان نیز نتیجه‌گیری و بحث مختصری پیرامون موضوع ارائه می‌شود.

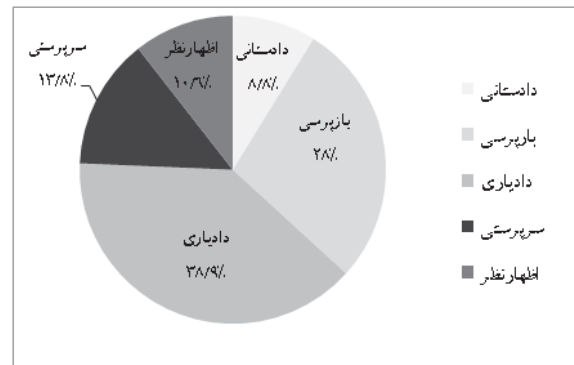
چنانکه در جدول ۱ اشاره شد، برای متغیرهای کیفی می‌توان از هر دو نمودار دایره‌ای و میله‌ای استفاده کرد. البته نمودارهای مدرن‌تری هم در این زمینه معرفی شده است و ما فقط به این دو نمودار که متداول شده می‌پردازیم. اما همانطور که پیش‌تر گفته شد مایلیم بدانیم آیا در عمل تمایزی بین این دو قایل شویم یا خیر. به زبان دیگر در چه مواردی بهتر است از نمودار میله‌ای استفاده شود و در چه مواردی نمودار دایره‌ای ارجحیت دارد. البته پاسخ به این سوال تا اندازه‌ای به سلیقه‌ی کاربران ارتباط دارد. اما با ذکر چند مثال نشان می‌دهیم که در برخی موارد می‌توان بین این دو یکی را بر دیگری ترجیح داد. شکل ۱، نمودار دایره‌ای مربوط به یکی از گزارش‌های تولید شده در یکی از بخش‌های سازمان قضایی است. در این شکل چند نکته وجود دارد که در ادامه مورد بحث قرار می‌دهیم. اولین نکته‌ای که از این شکل برمی‌آید آن است که متغیر مورد بحث (نوع شعبه‌ی دادسرا) کیفی است و لذا مطابق با جدول ۱ نمودار دایره‌ای می‌تواند انتخاب مناسبی باشد. اما در مورد ظاهر نمودار چند نکته‌ی جالب وجود دارد:

- برای تشخیص تفاوت بین سطوح متغیر، هم از رنگ‌های مختلف استفاده شده است و هم روی هر قطاع نام آن سطح نوشته شده است. واضح است که یکی از این دو اضافه است و می‌توان آن را حذف کرد.

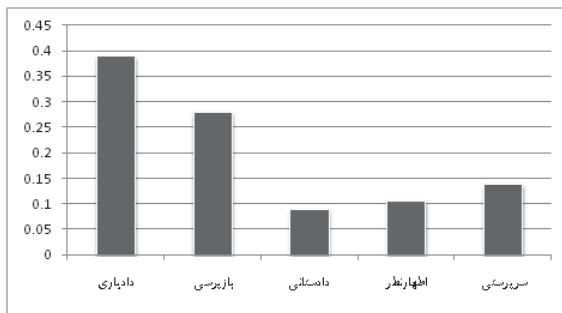
- برای تشخیص تفاوت مقادیر هر سطح اعداد مربوطه به عنوان برچسب به نمودار اضافه شده است. در اینجا با توجه به نزدیک بودن مقدار مربوط به



شکل ۲: توزیع فراوانی نسبی پرونده‌های ارجاع شده در شعب مختلف دادسراها.



شکل ۱: توزیع فراوانی نسبی پرونده‌های ارجاع شده در شعب مختلف دادسراها.



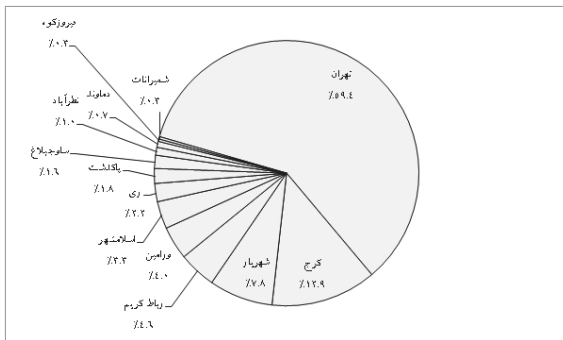
شکل ۳: توزیع فراوانی نسبی پرونده‌های ارجاع شده در شعب مختلف دادسراها.

دو سطح دادستانی و اظهار نظر، درج این اعداد اجتناب‌ناپذیر است. در واقع در صورتی که اعداد مربوط به دو سطح دادستانی و اظهار نظر روی نمودار درج نشده باشد، تشخیص تفاوت آن‌ها دشوار است. در حالیکه در صورت تفاوت زیاد بین قطعات نیاز به چنین امری وجود ندارد.

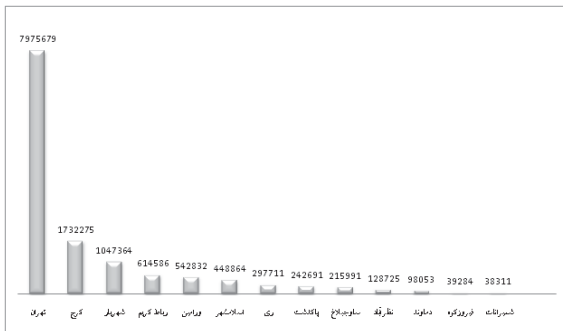
می‌توان به جای شکل ۱ از نمودار شکل ۲ استفاده کرد. در این صورت برخی از مواردی که در بالا ذکر شد، برطرف می‌شود. اما همچنان بدون داشتن مقادیر برجسب‌ها امکان مقایسه برای دو سطح دادستانی و اظهار نظر با توجه به نزدیکی مقادیر وجود ندارد. این بدان معناست که این نمودار بدون برجسب‌های عددی امکان مقایسه‌ی- کامل را فراهم نمی‌کند. حال به جای این دو نمودار از یک نمودار میله‌ای کمک می‌گیریم. در این صورت، نموداری مانند شکل ۳ به دست می‌آید. ملاحظه می‌شود که در این نمودار مشکلات قبلی کمتر به چشم می‌خورد؛ همانطور که ملاحظه می‌کنیم بدون نیاز به داشتن مقادیر هر سطح می‌توانیم آن‌ها را با یکدیگر مقایسه کنیم.

نمودار دایره‌ای نسبت به نمودار میله‌ای ارجحیت دارد. به عنوان نمونه شکل ۴ مربوط به سهم چند کشور دنیا از بازار بیمه‌ی غیر عمر برحسب درآمد حق بیمه‌ی آن‌هاست (به نقل از سالنامه‌ی آمارهای بیمه‌ی منتشر شده توسط سازمان همکاری و توسعه اقتصادی (OECD) در سال ۲۰۰۸). در این مثال اگر هدف مقایسه‌ی کشورها باشد به نظر نمودار میله‌ای بهتر است چون تفاوت‌های کمی مانند ژاپن و فرانسه در آن آسان‌تر قابل مشاهده است؛ اما در صورتی که بخواهیم سهم هر کشور را به صورت جداگانه از کل بازار برآورد کنیم، نمودار دایره‌ای کاراتر است. در واقع بدون داشتن درصدها به عنوان مثال بر اساس نمودار

البته شایان ذکر است که اگر هدف برآورد نسبت و سهم هر سطح از متغیر مورد بررسی باشد، آنگاه در برخی موارد



شکل ۵: توزیع فراوانی جمعیت در شهرهای مختلف استان تهران در سال ۱۳۸۵.



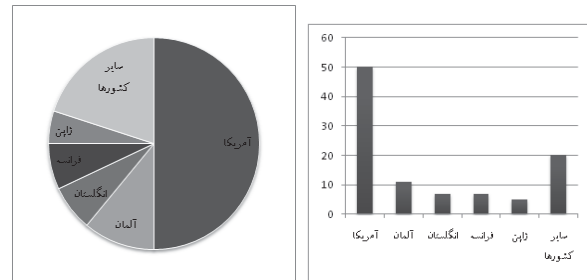
شکل ۶: توزیع فراوانی جمعیت در شهرهای مختلف استان تهران در سال ۱۳۸۵.

موارد مطرح شده به سادگی امکان‌پذیر است. اما با این وجود دو مشکل در نمایش این نمودار به چشم می‌خورد: مورد اول اینکه مانند مثال‌های قبل مقادیر مربوط به چند سطح بسیار به هم نزدیک بوده و تفاوت‌ها مشخص نیست؛ مورد دیگر مربوط به برچسب‌های محور افقی است که به دلیل طولانی بودن عناوین در چند سطر نمایش داده شده است. برای حل این دو مشکل پیشنهاد می‌کنیم که از قالب شکل ۸ استفاده شود. چنانکه ملاحظه می‌کنیم این نمودار نسبت به نمودار شکل ۷ زیباتر است و برای سطوح بیشتر از این مثال نیز قابل استفاده است.

مثال‌های بالا نشان می‌دهند که:

- در مواردی که تفاوت بین سطوح زیاد نباشد و هدف

دایره‌ای می‌توان گفت که آمریکا به تنهایی ۵۰ درصد کل بازار را در اختیار دارد. اما با استفاده از نمودار میله‌ای چنین نتیجه‌ای نمی‌توان گرفت.

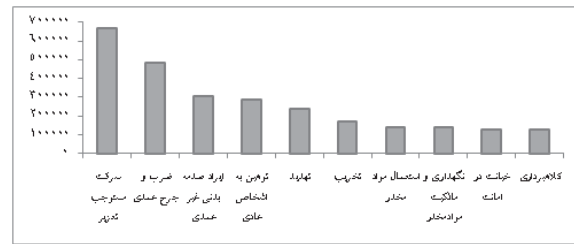


شکل ۴: سهم کشورهای مختلف از بازار بیمه‌ی غیر عمر در سال ۲۰۰۸.

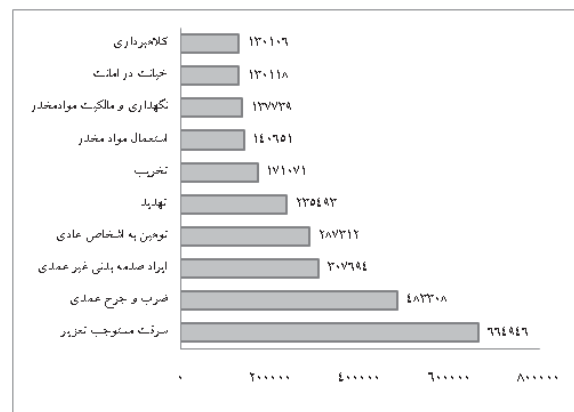
اکنون اجازه دهید مثال دیگری از مقایسه‌ی این دو نمودار ارائه کنیم. جمعیت شهرستان‌های استان تهران بر اساس سرشماری سال ۸۵ از وبگاه مرکز آمار ایران [۲] در دسترس است. با استفاده از این اطلاعات نمودار دایره‌ای و میله‌ای به ترتیب در شکل ۵ و شکل ۶ رسم شده است. همانطور که ملاحظه می‌شود وضوح هر دو نمودار در یک سطح است و برای نمودار میله‌ای نیز تمایز بین دو شهرستان شمیرانات و فیروزکوه بدون وجود برچسب مقادیر برای هر میله امکان‌پذیر نیست. در چنین شرایطی به نظر می‌رسد در مقایسه‌ی دو نمودار دایره‌ای و میله‌ای تفاوتی بین آن‌ها وجود ندارد. البته این مثال نشان می‌دهد که در صورت زیادتر بودن تعداد سطوح متغیر مورد بررسی استفاده از نمودار میله‌ای بر نمودار دایره‌ای ارجحیت دارد، چرا که نوشتن برچسب‌ها با افزایش تعداد آن‌ها در نمودار دایره‌ای به مراتب دشوارتر می‌شود.

به عنوان مثالی دیگر فرض کنید تعداد پرونده‌های مرتبط با چند عنوان اتهامی مانند نمودار شکل ۷ باشد. این نمودار وضوح خوبی دارد و امکان مقایسه‌ی انواع

۱۸۰۱ میلادی تا سال‌های اخیر بحث‌های زیادی در نقد و تایید نمودار دایره‌ای در مقابل نمودار میله‌ای مطرح شده است. اسپنس [۱۳] مرور نسبتاً جامعی بر این موضوع تا سال ۲۰۰۵ انجام داده است که البته در مجموع از نظر ایشان نباید از نمودار دایره‌ای بیش از حد توانش انتظار داشت. بنابر تحقیق ایشان برخی از محققین موافق استفاده از نمودار دایره‌ای و برخی مخالف آن بوده‌اند. نکته‌ی جالب در نمودار دایره‌ای آن است که پلی‌فیر که مبدع این نمودار بوده است در خصوص مزایا و معایب آن هیچ‌گونه مطلبی ننوشته و در ضمن نسبت به سایر نمودارها نیز کمتر از این نمودار استفاده کرده است. علاوه بر این اسپنس [۱۳] در تحقیق خود نمودارهای به کار رفته در چند مجله معروف را مورد بررسی قرار داده و نتیجه گرفته است که گرایش محققین به استفاده از نمودار دایره‌ای کمتر از سایر نمودارهاست. با این وجود تحقیق‌های انجام گرفته در این خصوص، نشان می‌دهد که دلیل محکمی برای حذف این نمودار وجود ندارد.



شکل ۷: تعداد پرونده‌های مرتبط با ده عنوان اتهامی اول کشور در یک ماه.



شکل ۸: تعداد پرونده‌های مرتبط با ده عنوان اتهامی اول کشور در یک ماه.

۳ نمودارهای آماری برای داده‌های کمی

پس از معرفی نمودارهای دایره‌ای و میله‌ای توسط پلی‌فیر، در سال ۱۸۲۱ فوریر برای جمعیت شهر پاریس از نمودار میله‌ای استفاده کرد و با در نظر گرفتن گروه‌های سنی و وصل کردن فراوانی آن‌ها با یک خط به همدیگر یک منحنی به دست آورد که بعدها گالتن در سال ۱۸۷۵ آن را نمودار اجایو^۱ نامید (بنیگر و رابین [۴]). علاوه بر این گوری نیز در سال ۱۸۳۳ داده‌های مربوط به جرایم را بر

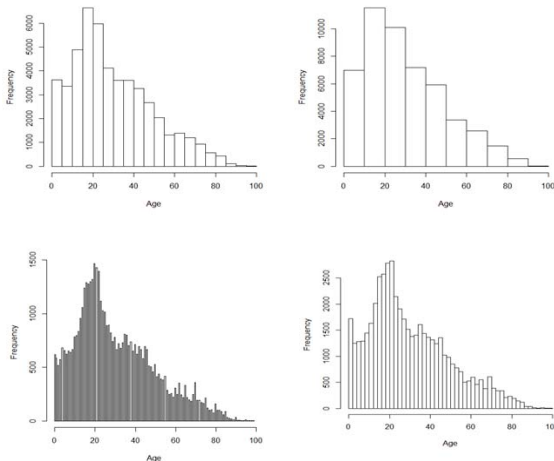
مقایسه‌ی سطوح باشد بهتر است از نمودار میله‌ای استفاده شود و در صورتی که هدف برآورد نسبت یا سهم هر سطح از کل باشد از نمودار دایره‌ای استفاده شود.

- در مواردی که تعداد سطوح زیاد باشد بهتر است از نمودار میله‌ای استفاده شود.
- در مواردی که برچسب‌های سطوح طولانی باشند یا تعداد سطوح زیاد باشند (به عنوان مثال استان) بهتر است از نمایش افقی نمودار میله‌ای استفاده شود.

موضوع استفاده از نمودار دایره‌ای یا نمودار میله‌ای از منظر محققین آماری دور نبوده و شاید پس از معرفی نمودار دایره‌ای به شکل امروزی توسط پلی‌فیر در سال

^۱Ogive

همچنین در نمودار دومی با ۲۰ رده این دسته‌بندی تقریباً ملموس است در حالی که برای نمودار اولی با ۱۰ رده چنین برداشتی به سختی وجود دارد.



شکل ۹: توزیع فراوانی سنی جمعیت شهرستان تویسرکان در سال ۱۳۸۵.

مثال بالا نشان می‌دهد که در برخی شرایط اضافه کردن تعداد رده‌ها در نمودار مستطیلی ممکن است نتایج دقیق‌تری فراهم کند. البته این موضوع بستگی به هدف بررسی نیز دارد. به عنوان مثال اگر هدف از بررسی تعیین نوع تقارن توزیع داده‌ها باشد، آنگاه در مثال بالا همان ۱۰ رده نیز کفایت می‌کند در حالی که برای تعیین نسل‌های جمعیتی بهتر است از تعداد رده‌های بیشتری استفاده شود. چنانکه پیش‌تر گفته شد، ابزارهای نموداری برای توصیف اطلاعات در یک دوره‌ی چند ده ساله دارای تحول چشم‌گیری نبوده و از آن طرف ظهور و گسترش تکنیک‌های آمار ریاضی برای مدل‌بندی داده‌ها به عنوان یک جایگزین گوی سبقت را از ابزارهای نموداری رده‌ای بوده. شاید اگر به تنها نمودار مناسبی که برای داده‌های کمی رایج شده بود (هیستوگرام) نگاه کنیم می‌بینیم که عدم

اساس رده‌بندی سنی با استفاده از نمودار میله‌ای ترسیم کرد که در واقع همان هیستوگرام امروزی است (بنیگر و رایین [۴]). البته واژه هیستوگرام در سال ۱۸۹۵ برای اولین بار توسط کارل پیرسن مورد استفاده قرار گرفت. یکی از مسائلی که در ترسیم نمودار هیستوگرام باید بدان توجه کرد، تعداد مستطیل‌ها یا همان رده‌ها است. این مسئله در واقع همان تعداد رده‌هایی است که برای تشکیل جدول فراوانی نیز مورد بحث قرار می‌گیرد. اگر چه در برخی کتب آماری فرمول‌هایی نیز برای این منظور ارائه می‌دهند اما به نظر می‌رسد که این فرمول‌ها کلیت نداشته و بلکه باید با توجه به نوع داده‌ها در این خصوص تصمیم‌گیری کرد. برخی از معروف‌ترین این فرمول‌ها مانند فرمول استورگس [۱۴]، فرمول اسکات [۱۲] و فرمول فریدمن و دیاکونیس [۷] معمولاً در نرم‌افزارهای آماری مورد استفاده قرار می‌گیرند. پژوهشگرانی که درباره‌ی مناسب بودن این فرمول‌ها تحقیق کرده‌اند به این نتیجه رسیده‌اند که این فرمول‌ها برای داده‌هایی که تعداد آن‌ها کم است (مثلاً کمتر از ۲۰۰) می‌تواند مناسب باشد اما برای اندازه‌های بزرگ‌تر به هیچ‌وجه مناسب نیستند (به عنوان مثال هاوتون و فون [۹] و هیندمن [۱۰] را ببینید).

اگرچه علاوه بر تعداد رده‌ها مقادیر داده‌ها نیز در تعیین رده‌ها نقش دارند. در واقع ممکن است بر اساس مقادیر داده‌ها تصمیم بگیریم که از رده‌های با طول برابر یا نابرابر استفاده کنیم. شکل ۹ نمودار هیستوگرام مربوط به جمعیت شهری شهرستان تویسرکان را بر اساس نتایج سرشماری سال ۱۳۸۵ با در نظر گرفتن تعداد رده‌های مختلف نشان می‌دهد. در این نمودارها به ترتیب ۱۰، ۲۰، ۵۰ و ۱۰۰ رده در نظر گرفته شده است. ملاحظه می‌کنیم که دو نمودار آخری با ۵۰ و ۱۰۰ رده تصویر واضحی از حضور سه نسل جوان، میانسال و کهنسال ارائه می‌کند؛

۲	۱۳۶
۳	۴۸۸
۴	۰۱۱۲۴۵۸۹
۵	۰۰۰۰۱۴۶۸۹
۶	۰۱۲۴۷
۷	۹
۸	۵

شکل ۱۰: توزیع فراوانی متوسط زمان بررسی پرونده‌های قضایی (به روز).

مربوط به نحوه‌ی تفسیر این نمودار است تا نحوه‌ی ترسیم آن. آن‌ها معتقدند که اضافه کردن توضیح‌هایی بر این نمودار و انجام تغییرات کوچکی در ظاهر آن می‌تواند به کم شدن این مشکلات کمک کند. از این جهت ویرایش‌های مختلفی نیز توسط آن‌ها و محققان دیگری معرفی شده است که امروزه در گزارش‌ها از آن‌ها استفاده می‌شود (برای مثال توکی [۱۵]، مک گیل و همکاران [۱۱] و گلدبرگ و ایگلوویچ [۸] را ببینید). قابل ذکر است که این نمودار علاوه بر جنبه‌ی ظاهری قابل مقایسه‌ی آن با دو نمودار هیستوگرام و نمودار ساقه و برگ، چند شاخص عددی را نیز به همراه دارد که در تفسیر نمودار استفاده‌ی مفیدی دارد: این شاخص‌ها شامل کران‌های پایین و بالا، چارک‌های اول، دوم و سوم است که به صورت مستقیم بر این نمودار نمایش داده می‌شوند. از کاربردهای متمایز این نمودار تعیین نقاط دورافتاده است که شاید با تغییر تعداد مستطیل‌ها در نمودار مستطیلی قابل رویت نباشد. البته در این خصوص باید توجه داشت که نتیجه‌گیری و گزارش در مورد دورافتاده بودن یک مشاهده باید با

اجماع نظر در انتخاب تعداد رده‌ها می‌تواند به عنوان یکی از دلایل عدم استقبال از این نمودارها باشد. برای این منظور توکی [۱۵] نمودار ساقه و برگ را به عنوان یک ابزار شبیه به نمودار هیستوگرام معرفی کرد که ضمن رسم نمودار امکان مشاهده‌ی خود مقادیر داده‌ها نیز بر نمودار فراهم است. در چنین نموداری ارزش مکانی ارقام هر عدد است که رده‌ها یا همان ساقه‌ها را مشخص می‌کند. مثالی از چنین نموداری برای متوسط زمان رسیدگی به پرونده‌های قضایی برحسب روز در شکل ۱۰ نشان داده شده است. در این نمودار دهگان را به عنوان ساقه و یکان هر عدد را به عنوان برگ انتخاب کرده‌ایم. این موضوع تا حدود زیادی مشکل رده‌بندی را حل می‌کند؛ ضمن اینکه همه‌ی داده‌های خام نیز در نمودار وجود دارد که یک حسن بزرگ محسوب می‌شود. اما بزرگ‌ترین عیبی که بر این نمودار وارد است مربوط به گستره‌ی کاربرد آن است: چنانچه تعداد داده‌ها از یک حدی بیشتر باشد امکان ترسیم این نمودار بسیار دشوار خواهد بود و کارایی خود را از دست می‌دهد. در چنین شرایطی مجدداً نمودار هیستوگرام به عنوان جایگزین مورد استفاده قرار می‌گیرد. تحلیل بالا نشان می‌دهد که نمودار ساقه و برگ می‌تواند زمانی که تعداد مشاهده‌ها خیلی زیاد نیست به عنوان یک ابزار مطلوب مورد استفاده قرار بگیرد. البته شایان ذکر است که در چنین نموداری زیبایی ظاهری نسبت به هیستوگرام بسیار پایین است و از این جهت است که شاید کمتر از این نمودار در گزارش‌های آماری استفاده می‌شود. از دیگر نمودارهایی که به منظور تقویت توصیف داده‌های کمی معرفی شده است نمودار جعبه‌ای است. این نمودار نیز توسط توکی [۱۵] در حدود ۱۹۷۰ معرفی شده است. بنا به گفته‌ی مک‌گیل و همکاران [۱۱] ضعف‌هایی نیز برای این نمودار وجود دارد که بیشتر آن‌ها

احتیاط صورت بپذیرد. در تحقیقی که داوسن [۶] بر اساس شبیه‌سازی انجام داده است نشان داده است که در نمونه‌های با اندازه‌ی کوچک (کمتر از ۲۱) از توزیع نرمال بیش از ۳۰ درصد مشاهده‌ها بر اساس رویه‌ی نمودار جعبه‌ای به عنوان مشاهده‌ی دورافتاده نمایش داده می‌شوند؛ در حالیکه باید این مقدار از لحاظ نظری کمتر از یک درصد باشد. بنابراین در صورتی که از نمودار جعبه‌ای برای داده‌های با تعداد کم استفاده شد و داده‌ای در نمودار به صورت دورافتاده مشاهده شد، نمی‌توان با قطعیت آن را به عنوان مشاهده‌ی دورافتاده معرفی کرد. در مجموع ملاحظه می‌کنیم که نمودارهای کمی از لحاظ ظاهری گسترش چندانی نداشته‌اند و نمودارهای ساقه و برگ و نمودار جعبه‌ای نمودارهای فنی‌تری نسبت به هیستوگرام محسوب می‌شوند که ترسیم و تفسیر آن‌ها نیز به مراتب دشوارتر است.

۴ نمودارهای چند متغیری

در تحقیق‌های عملی معمولاً بیش از یک متغیر مورد بررسی قرار می‌گیرد. بنابراین برای توصیف مشاهده‌های به دست آمده باید راهکار مناسبی انتخاب کرد. در ساده‌ترین حالت اگر دو متغیر در اختیار باشد، برای برخی از انواع متغیرها می‌توان از همان نمودارهای یک متغیره با آرایش خاصی استفاده کرد. به عنوان مثال اگر در مثال مربوط به جمعیت شهرهای استان تهران علاوه بر نام شهر، جنسیت افراد نیز در نظر گرفته شود آنگاه می‌توانیم از همان نمودار میله‌ای اما با در نظر گرفتن دو میله برای هر شهر (یک میله برای جمعیت زنان و دیگری برای جمعیت مردان) استفاده کنیم. همچنین اگر در مثال مربوط به متوسط زمان رسیدگی دو نوع پرونده (مثلاً کیفری و

حقوقی) را در نظر بگیریم، می‌توان از دو نمودار جعبه‌ای در کنار هم استفاده کرد. اما در برخی شرایط استفاده از نمودارهای دیگری مانند نمودارهای خطی و پراکنش ارجحیت دارد. به عنوان مثال اگر بخواهیم ارتباط بین دو متغیر کمی را بررسی کنیم استفاده از نمودار پراکندگی گزینه مناسبی خواهد بود. چرا که با ترسیم این نمودار به سادگی می‌توان در مورد نوع ارتباط و حتی تا حدودی شدت ارتباط متغیرهای مورد بررسی اظهار نظر کرد. همچنین در صورتی که مشاهده‌هایی از نوع سری زمانی در اختیار باشد با رسم نمودار خطی به سادگی می‌توان در مورد روند تغییر سری و مولفه‌های دیگری مانند مولفه‌های دوره‌ای آگاهی‌های مناسبی به دست آورد. استفاده از این نمودارها به ویژه برای تسریع در مقایسه جوامع بسیار مفید هستند. به عنوان مثال در جدول ۲ مشاهده‌های مربوط به دوازده ماه برای آمریکا و اروپا گزارش شده است. مطابق با این جدول برای مقایسه دو جامعه باید ماه به ماه مشاهده‌ها را مقایسه کرده و در ضمن در مورد روند تغییر نیز به سختی می‌توان صحبت کرد. اما نمایش دیگری از این داده‌ها با استفاده از نمودار خطی در شکل ۱۱ ارائه شده است. بر اساس این شکل به سادگی امکان مقایسه این دو و حتی روند تغییر در هر یک به طور جداگانه به سادگی قابل بررسی است.

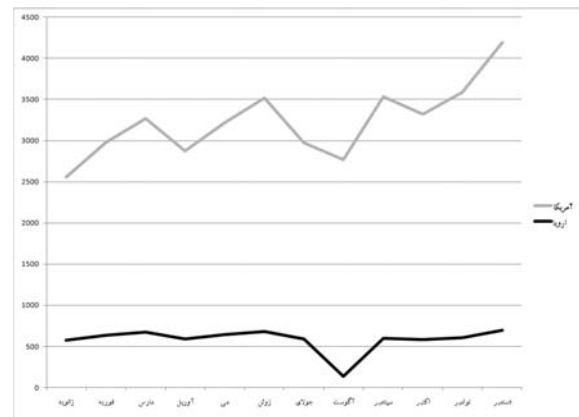
۵ نتیجه‌گیری

مطالعه‌ی سیر تاریخی به وجود آمدن نمودارهایی که در آمار توصیفی رایج هستند بر اساس سوابق موجود در منابع علمی و اطلاعاتی یکی از دستاوردهای این مقاله است. در این میان نقش محققین مختلف در به وجود آمدن و گسترش این ابزارهای توانمند و البته فراز و

جدول ۲: حجم تماس‌های تلفنی (به هزار) به تفکیک ماه در اروپا و آمریکا در سال ۲۰۰۳

ماه	ژانویه	فوریه	مارس	آوریل	می	ژوئن
آمریکا	۱۹۸۳	۲۳۴۳	۲۵۹۳	۲۲۸۳	۲۵۷۴	۲۸۳۸
اروپا	۵۷۴	۶۳۶	۶۷۳	۵۹۳	۶۴۴	۶۷۹
ماه	جولای	آگوست	سپتامبر	اکتبر	نوامبر	دسامبر
آمریکا	۲۳۸۲	۲۶۳۴	۲۹۳۸	۲۷۳۹	۲۹۸۳	۳۴۹۳
اروپا	۵۹۳	۱۳۹	۵۹۹	۵۸۳	۶۰۲	۶۹۰

در عمل با توجه به نوع داده‌ها انتخاب رده‌ها باید صورت بگیرد. همچنین توجه به جنبه‌ی زیبایی نمودارها و سادگی آن‌ها از دیگر مواردی است که نگارنده‌گان تأکید به توجه به آن‌ها در استفاده از نمودارها دارند.



مراجع

شکل ۱۱: حجم تماس‌های تلفنی در سال ۲۰۰۳ در آمریکا و اروپا [۱] بهبودیان، ج. (۱۳۸۰). آمار و احتمال مقدماتی. انتشارات آستان قدس رضوی. چاپ چهاردهم.

[۲] درگاه ملی آمار:

URL: <http://www.amar.org.ir>.

[3] Anscombe, F. J. (1973). Graphs in statistical analysis. *The American Statistician*, 27(1), 17-21.

[4] Beniger, J. R. and Robyn, D. L. (1978). Quantitative graphics in statistics: a brief history. *The American Statistician*, 32 (1), 1-11.

فرودهای آن نیز مورد بحث قرار گرفت. به طور خاص در مورد نمودارهای دایره‌ای، میله‌ای، هیستوگرام، ساقه و برگ و جعبه‌ای به عنوان نمودارهای رایج مطالب کاربردی و مفیدی به همراه چند مثال ارائه شد. بر این اساس نکاتی را که در به کار گرفتن این نمودارها باید به آن‌ها توجه کرد مورد بحث قرار دادیم و دیدیم که عدم توجه به این نکات می‌تواند در برخی موارد گمراه‌کننده باشد یا در شرایط بهتر از کیفیت گزارش بکاهد. نتایج حاصل از این مقاله نشان می‌دهد که برخی از بحث‌های مرتبط با این موضوع، مانند نحوه‌ی رده‌بندی در نمودارهای هیستوگرام، هنوز جزء مسائلی است که نیازمند توجه بیشتری هستند و

- [12] Scott, D.W. (1979). On optimal and data-based histograms. *Biometrika*, 66, 605-610.
- [13] Spence, I. (2005). No humble pie: the origins and usage of a statistical chart. *Journal of Educational and Behavioral Statistics*, 30(4), 353-368.
- [14] Sturges, H. (1926). The choice of a class-interval. *Journal of American Statistical Association*, 21, 65-66.
- [15] Tukey, J. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.
- [5] Cox, N. J. (2007). The grammar of graphic. *Journal of Statistical Software*. 17, Book review 3.
- [6] Dawson, R. (2011). How significant is a boxplot outlier?. *Journal of Statistics Education*, 2(1), 1-13.
- [7] Freedman, D. and Diaconis, P. (1981). On this histogram as a density estimator: L2 theory. *Zeit. Wahr. ver. Geb.*, 57, 453-476.
- [8] Goldberg, K.M. and Iglewicz, B. (1992). Bivariate extensions of the boxplot. *Technometrics*, 34, 307-320.
- [9] Houghton, N. and Phong, N. (2004). Graphical and numerical descriptive analysis: exploratory tools applied to vietnamese data. *Journal of Statistics Education*, 12(2).
- [10] Hyndman, R. (1995). The problem with sturges' rule for constructing histogram. Available at: <http://www.robjhyndman.com/papers/sturges.pdf>.
- [11] McGill, R., Tukey, J.W. and Larsen, W.A. (1978). Variations of boxplots. *The American Statistician*, 32(1), 12-16.