



پیام نوریها
public channel



✓ کانال پیام نوریها در سال 95 با هدف تهیه جزوات و نمونه سوالات افتتاح و از همان ابتدای تاسیس کوشیده است با تکیه بر تلاش بی وقفه، کارگروهی و فعالیت های بدون چشمداشت کاربران متمایز خود، قدمی کوچک در راه پیشرفت ارائه خدمات به دانشجویان این مرز و بوم بردارد.

@Payamnoria

[telegram.me/Payamnoria](https://t.me/Payamnoria)

رایگان است و همیشه رایگان میماند



اطلاع از اخبار و دانلود جزوات و نمونه سوالات

[برای ورود به کانال تلگرامی پیام نوریها کلیک کنید](#)

" کانال و خانواده تلگرامی پیام نوریها "

با عضویت در کانال و به آرشیو زیر دسترسی پیدا کنید

✓ تمام نمونه سوالات به روز تا آخرین دوره

✓ جزوات درسی

✓ بیش از ۱۰۰ فلش کارت دروس

✓ اخبار به روز پیام نور

✓ فیلم و فایل آموزشی اختصاصی

✓ انجام انتخاب واحد و حذف و اضافه

✓ پاسخگویی به سوالات دانشجویان

✓ معرفی گروه و انجمن های پیام نوری

✓ طنز و توییت دانشجویی

به یکی از بزرگترین کانال های پیام نوری بپیوندید

[برای ورود به کانال تلگرامی پیام نوریها کلیک کنید](#)

جزوه مبانی احتمال

نویسندگان : دکتر نرگس عباسی ، دکتر علی شادرخ و دکتر محمد قاسم وحیدی اصل

فهرست

ط	پیشگفتار
۱	فصل اوّل: آمار توصیفی
۲	مقدمه
۳	۱-۱ اعداد
۵	۲-۱ آمار توصیفی و آمار استنباطی
۶	۳-۱ برخی مفاهیم اوّلیه
۸	۴-۱ مفهوم مقیاس سازی
۱۱	۵-۱ انواع داده ها
۱۳	تمرین
	خلاصه ی فصل اول
۱۴	خودآزمایی چهارگزینه ای
۱۷	فصل دوّم جدول و نمودارهای آماری
۱۸	مقدمه
۱۹	۱-۲ مرتب کردن داده های عددی
۲۲	۲-۲ توزیع های فراوانی
۲۸	۳-۲ توصیف داده ها به کمک نمودارها
۳۲	خلاصه ی فصل دوم
۳۲	تمرین
۳۴	خودآزمایی چهارگزینه ای
۳۷	فصل سوّم معیارهای تمرکز و پراکندگی
۳۸	مقدمه
۳۸	۱-۳ معیارهای گرایش مرکزی

۳۸	۱-۱-۳ میانگین
۴۱	۲-۱-۳ میانه
۴۶	۳-۱-۳ مد
۴۷	۴-۱-۳ دیگر معیارهای گرایش مرکزی
۵۰	۲-۳ معیارهای پراکندگی
۵۱	۱-۲-۳ دامنه
۵۲	۲-۲-۳ انحراف معیار
۵۸	۳-۲-۳ فرمول‌های دیگر برای محاسبه‌ی پراکندگی
۵۹	۳-۳ کاربردهایی از معیارهای پراکندگی
۵۹	۱-۳-۳ ضریب تغییر
۶۰	۲-۳-۳ نمره‌ی استاندارد
۶۰	۳-۳-۳ قضیه چبیشف
۶۲	۴-۳ چندک‌ها و دیگر معیارهای پراکندگی
۶۶	۵-۳ نمودار جعبه‌ای
۶۸	خلاصه‌ی فصل سوم
۷۰	تمرین
۷۱	خودآزمایی چهارگزینه‌ای
۷۵	فصل چهارم معیارهای ارزیابی منحنی فراوانی
۷۶	مقدمه
۷۶	۱-۴ معیارهای ویژه‌ی نمودارهای داده‌های
۸۲	۲-۴ نمودار جعبه - خط و ارزیابی فراوانی‌ها
۸۴	خلاصه‌ی فصل چهارم
۸۴	تمرین
۸۵	خودآزمایی چهارگزینه‌ای
۸۷	فصل پنجم روش‌های اساسی شمارش
۸۸	مقدمه
۸۸	۱-۵ اصول شمارش
۹۰	۲-۵ جایگشت
۹۱	۳-۵ ترتیب
۹۲	۴-۵ جایگشت دوری
۹۳	۵-۵ ترکیب
۹۶	۶-۵ تعمیم ترکیب
۹۷	۷-۵ ضرایب دوجمله‌ای

۱۰۰	خلاصه‌ی فصل پنجم
۱۰۱	تمرین
۱۰۴	خودآزمایی چهارگزینه‌ای
۱۰۹	فصل ششم احتمال
۱۱۰	مقدمه
۱۱۰	۱-۶ آزمایش تصادفی
۱۱۱	۲-۶ فضای نمونه‌ای و برآمد
۱۱۳	۳-۶ پیشامد
۱۱۶	۴-۶ احتمال
۱۲۲	۵-۶ برخی قواعد احتمال
۱۲۷	۶-۶ احتمال شرطی
۱۳۳	۷-۶ پیشامدهای مستقل
۱۳۶	۸-۶ قضیه بیز
۱۴۰	خلاصه‌ی فصل ششم
۱۴۱	تمرین
۱۴۹	خودآزمایی چهارگزینه‌ای
۱۵۳	فصل هفتم محاسبات با نرم‌افزار
۱۵۴	مقدمه
۱۵۴	۱-۷ نحوه‌ی ورود اطلاعات
۱۵۶	۲-۷ ساخت جدول آماری و رسم نمودارها
۱۶۳	۳-۷ محاسبه‌ی معیارهای مرکزی و پراکندگی
۱۶۵	خلاصه‌ی فصل هفتم
۱۶۵	۴-۷ تمرین
۱۶۷	پاسخ تمرین‌ها
۱۸۶	پاسخنامه خودآزمایی‌های چهارگزینه‌ای
۱۸۷	مرجع‌ها

پیشگفتار

اکثر تحقیقات امروز نیازمند علم آمار است و این کتاب می‌تواند شروعی خوب برای یادگیری آن باشد. این کتاب برای درس «مبانی احتمال» بر اساس سرفصل دروس رشته‌های «آمار و کاربردها» و «ریاضی و کاربردها» نوشته شده است. با پرهیز از آوردن مطالب تکراری، سعی شده است که مطالب ارائه شده به سادگی و بدون شاخه و برگ دادن، بیان شوند تا خواننده بدون نیاز به ریاضیات پیشرفته، به راحتی موضوع را درک نماید، و پایه‌ی خوبی برای گذراندن دروس تخصصی آماری بعدی داشته باشد.

از آنجاکه داده‌ها به صورت انبوه در مسائل کاربردی در دسترس هستند، آموزش یک نرم‌افزار آماری در سطح مقدماتی لازم به نظر می‌رسید از این رو نرم‌افزار «آر» را به دلیل سادگی و قابلیت‌های برتر آن نسبت به سایر نرم‌افزارها، برای آموزش در کتاب، انتخاب کردیم. دو بخش تمرینات و خودآزمایی چهارگزینه‌ای به همراه پاسخ برای ذکر برخی نکات و ویژگی‌ها و افزایش یادگیری، در کتاب قرار گرفته است. در مطالعه‌ی کتاب، خواننده می‌تواند فصل‌های پنجم و ششم، را ابتدا مطالعه نماید زیرا مطالب فصل‌های اول الی چهارم که به آمار توصیفی می‌پردازد، پیش‌نیاز آن‌ها نیستند.

تمام تلاش نویسندگان بر این بوده است که اثر حاضر هیچ‌گونه نقایص علمی و تاییپی نداشته باشد، باوجود این از کلیه‌ی همکاران و دانشجویانی که نظرات اصلاحی خود را به اطلاع برسانند سپاسگزار خواهیم بود.

در اینجا لازم می‌دانیم از جناب آقای دکتر عبدالرضا بازرگان، و جناب آقای دکتر علیرضا نعمت‌الهی، ویراستار محترم کتاب، به خاطر دقت و حوصله در مطالعه‌ی متن اولیه و به خاطر پیشنهادهای ارزنده‌شان سپاسگزاری نماییم. همچنین از شورای تخصصی گروه آمار، مدیریت محترم تدوین، و سرکار خانم معدلت کمال تشکر را دارم.

نویسندگان

۱۳۹۵/۰۴/۱

فصل اول

آمار توصیفی

هدف کلی

آشنایی با مفاهیم مورد نیاز جهت ورود به مبحث آمار توصیفی

هدف‌های یادگیری

- خواننده پس از یادگیری این فصل باید بتواند:
- مثال‌هایی تاریخی از کاربرد علم آمار بیان نماید.
- جامعه‌ی آماری و انواع آن را تعریف کند.
- مفهوم نمونه‌ی تصادفی را شرح دهد.
- مفهوم علم آمار را بیان کند.
- نقش اعداد در رویدادهای طبیعی و اندیشه‌های انسانی را شرح دهد.
- آمار توصیفی را تعریف کند.
- توضیح دهد که خلاصه‌سازی در آمار توصیفی به‌وسیله‌ی چه چیزهایی انجام می‌شود.
- آمار استنباطی را توضیح دهد.
- انواع متغیرها را با ذکر مثال‌هایی مناسب بیان نماید.
- انواع مقیاس‌ها و ویژگی‌های آنها را با ذکر مثال‌هایی مناسب شرح دهد.
- مفهوم داده‌های خام را با ذکر مثال توضیح دهد.
- مفهوم داده‌های ثانویه را با ذکر مثال توضیح دهد.

مقدمه

کلمه‌ی آمار به دو معنی به کار می‌رود؛

معنای رایج آن معرف مجموعه‌ای از داده‌های است که دامنه‌ی وسیعی از اشیاء، نظیر اندازه‌ی جوامع، اطلاعات فعالیت‌های تولیدی، قیمت اجناس، درآمدها، میزان بارندگی و غیره را شامل می‌شود. اطلاعات آماری از این نوع به طور متعارف توسط دولت، بخش‌های خصوصی و قسمت‌هایی از زندگی روزمره تولید می‌شود. خلاصه آنکه این کلمه به نظریه و روش‌های جمع‌آوری، توصیف و تحلیل داده‌های عددی مربوط می‌شود. به بیانی دیگر، آمار را باید علم و عمل استخراج، بسط، و توسعه‌ی دانش‌های تجربی انسانی با استفاده از روش‌های گردآوری، تنظیم، پردازش و تحلیل داده‌های تجربی (حاصل از اندازه‌گیری و آزمایش) دانست.

از نظر تاریخی می‌توان گفت از لحظه‌ای که شمارش اختراع شد، علم آمار نیز گسترش یافت. در زمان‌های قدیم دولت‌ها اطلاعات مربوط به جامعه، ثروت، تعداد سربازها و ... را برای کنترل اوضاع کشور و جمع‌آوری مالیات، گردآوری می‌کردند. به لحاظ تاریخی، چینی‌ها، چند هزار سال قبل از تولد مسیح، تعداد نفوس را برآورد کرده‌اند و مصریان نیز علاوه بر سرشماری نفوس، جداول تحرک جامعه را نیز تشکیل داده بودند. در ایران و هند نیز جمع‌آوری آمارهای اداری بیش از دوهزار سال سابقه دارد و چنین آورده‌اند که در زمان داریوش پادشاه ایران برای شمارش سربازان بدین‌گونه عمل می‌شد که آنان از جایگاه خاصی بگذرند و پاره‌سنگی در آن بگذارند. (جان‌فروند، ۱۳۷۵: ۵)

در قرن هفدهم آمار حیاتی توسط جان‌گرانث انگلیسی پایه‌گذاری شد. او را پدر علم آمار حیاتی می‌دانند که برای اولین بار آمارهای تولد و مرگ را بررسی کرد و جداول مرگ‌ومیر را تنظیم و امید به زندگی در سنین مختلف را محاسبه کرد و سرانجام در اواخر قرن هفدهم اولین مؤسسه‌ی بیمه‌ی عمر در لندن پا به عرصه وجود نهاد.

نظریه‌ی آمار از نیمه‌ی قرن هفدهم با وارد شدن نظریه‌ی احتمال و مسائل بازی‌های شانسی توسعه یافت. امروزه «رونالد فیشر» انگلیسی را به عنوان پدر علم آمار می‌شناسند. وی آمار را در زمینه‌های مختلف مانند ژنتیک، بیومتری، آموزش و پرورش، کشاورزی و مانند آن‌ها وارد کرد و در معرفی استنباط آماری شامل برآوردهای نقطه‌ای،

۳ مبانی آمار توصیفی

توزیع‌های دقیق نمونه، و طرح آزمایش‌ها پیش‌گام بود. بخش مهم دیگری از علم آمار، که «زمین‌آمار» نامیده می‌شود، در آن به مختصات داده‌های مربوط به جامعه‌ی تحت بررسی، و به تبع آن ساختار فضایی داده‌ها، مورد مطالعه قرار می‌گیرد. (محمدزاده: ۱۳۹۱)

به کارگیری روش‌های آماری در علوم اجتماعی و رفتاری کمی قبل از جنگ جهانی دوم شروع شد. تعداد آمارگیری‌ها در زمینه‌های مختلف افزایش یافت و ضرورت تفسیر اطلاعات مربوط به روان‌شناسی و تعلیم و تربیت آشنایی با علم آمار را اجتناب‌ناپذیر کرد. (ساروخانی، ۱۳۷۲ و نایی، ۱۳۷۶) امروزه موفقیت در بسیاری از زمینه‌های علمی مانند علوم انسانی، پزشکی و فنی و مهندسی بدون داشتن اطلاعات لازم از علم آمار مشکل و گاهی غیرممکن است.

۱-۱ اعداد

به کاربردن مفهوم عدد تحولی در بیان اندیشه‌های انسانی ایجاد کرده است. هم‌چنان‌که گذشتگان دور نمی‌توانستند، رویدادهای طبیعی را به اندیشه و گفتار درآورند، وضع بشر، تا پیش از اختراع دماسنج، برای بیان وضع هوا مانند گذشتگان دور خود بود. اگر از وی وضع هوا را از لحاظ گرمی و سردی می‌پرسیدند، مثلاً جواب می‌داد: هوا سرد است. اگر می‌گفتند چقدر سرد است، می‌گفت: «خیلی» سرد است. اما با اختراع دماسنج و کمی‌شدن یا به کمیت درآمدن میزان دما، بشر قادر است به آسانی مناطق «سردسیر» و «گرمسیر» و مانند آنها را به صورتی مشخص‌تر تعریف کند و حتی رده‌بندی را بسیار دقیق‌تر و گسترده‌تر کند. بیان اینکه فلان کشور «پرجمعیت» است، نمی‌تواند در این حد از دقت باشد که بگوییم مثلاً جامعه آن کشور یک «میلیون نفر» است، یا متوسط تعداد کسانی که در هر هکتار زندگی می‌کنند مثلاً ۳۰۰۰ نفر است.

پس، اگر نتوانیم بگوییم که مفهوم عدد، زندگی انسان را دگرگون کرده است، لااقل می‌توانیم بگوییم که با متحول‌شدن زندگی انسان، عدد نقشی هرچه بهتر و بیشتر در زندگی او داشته است.

عدد در زندگی انسان نقش‌هایی گوناگون دارد. یونانیان حدود سه هزار سال پیش، عدد را تنها به‌خاطر خود عدد و برای درک روابط اعداد صحیح مورد مطالعه قرار دادند و زیباترین قضیه‌های ریاضی و موضوعی به‌نام «نظریه‌ی اعداد» را پدید آوردند. اما بشر،

در تمام عرصه‌های زندگی خود، چه شخصی و چه اجتماعی، به «حساب کردن» نیز نیاز داشت. این نیازهای زندگی، «چهار عمل اصلی حساب» و لذا «علم حساب» را پدید آورد که آن هم جزئی از ریاضیات است.

اما امروزه عدد نقشی بسیار فراتر از نقش خود در علم حساب دارد. در واقع بسیاری از پدیده‌های طبیعی و اجتماعی و مانند آنها را می‌توان به کمک اعداد بیان کرد و یا به تعبیر و تفسیر این پدیده‌ها به کمک اعداد پرداخت. مثلاً بیان شدت زلزله در واحد ریشتر، میزان قوت و ضعف نسبی آن را مشخص می‌کند، اما تعداد تلفات انسانی، تعداد بناهای تخریب شده و ... می‌تواند به بهترین نحو، ابعاد فاجعه را نشان دهد. پس در این مورد، چگونگی زلزله و عوارض و عواقب آن را با استفاده از اعداد توصیف کرده‌ایم. امروزه تلاش می‌شود که برای بیان، تجزیه و تحلیل، و تعبیر و تفسیر هر نوع پدیده‌ای از نیروی جادویی اعداد استفاده شود. البته نباید تصور کرد که بیان هر نوع پدیده‌ای و تجزیه و تحلیل آن با همان سهولتی انجام می‌شود که در بالا در مورد زلزله بیان شد. برای آنکه تصویری از برخی از این پدیده‌ها و پیچیدگی نسبی آنها داشته باشیم، به مثال‌های زیر توجه کنید:

- الف. تأثیر آموزش‌های پیش‌دبستانی در رشد رفتارهای اجتماعی،
- ب. نقش افزایش یا کاهش درآمد کشاورزی در مهاجرت‌های روستایی،
- ج. گستره‌ی شبکه‌ی روابط اجتماعی و تأثیر تحصیلات دانشگاهی،
- د. تأثیر آلودگی هوا در سطح سلامت جوامع،
- ه. تأثیر استفاده از وسایل فناوری پیشرفته در تغییر ساختار روابط اجتماعی.

برای تحلیل و بررسی اینگونه مثال‌ها، ابتدا باید اطلاعاتی عددی در زمینه‌ی موضوع مورد بحث گردآوری کنیم. چنین اطلاعاتی، برحسب مجموعه‌ای از اعداد، داده‌های آماری نامیده می‌شوند. نحوه‌ی گردآوری، تنظیم و تلخیص، تجزیه و تحلیل، و تعبیر و تفسیر داده‌ها، و استخراج نتایجی درباره‌ی موضوع مورد بحث، موضوع **علم آمار** را تشکیل می‌دهد. به بیانی دیگر، این علم شامل برنامه‌ریزی و جمع‌بندی و تفسیر مشاهدات غیرقطعی است به گونه‌ای که اعداد نماینده‌ی واقعی مشاهدات بوده و غیرواقعی یا نادرست نباشند و به نحو مفیدی تهیه و تنظیم، تحلیل و قابل نتیجه‌گیری صحیح باشند.

طبق نظر بهبودیان (۱۳۹۰) داده‌ها، یعنی اطلاعاتی هستند که به زبان اعداد ترجمه شده‌اند، اغلب دستخوش عوامل شانس می‌باشند و در صورتی می‌توانند روشنگر واقعیت‌ها باشند که:

الف. طبق قواعدی تهیه و تنظیم گردند.

ب. قوانین شانس که بر آنها حاکم می‌باشند، مشخص شوند.

ج. طبق اصولی با اطمینان کافی تجزیه و تحلیل شوند. (بهبودیان، ۱۳۹۰)

۱-۲ آمار توصیفی و آمار استنباطی

علم آمار بر دو شاخه‌ی اصلی آمار توصیفی و آمار استنباطی مبتنی است. در آمار توصیفی با داشتن تمام اعضای جامعه یا بخشی از آن، به توصیف ویژگی‌های آماری آن پرداخته می‌شود در حالی که در آمار استنباطی به کمک روش‌های آماری، استنباط‌هایی برای ویژگی‌های جامعه انجام می‌شود. در ادامه به توضیح بیشتر برای این دو موضوع، آمار توصیفی و آمار استنباطی، می‌پردازیم.

موضوع آمار توصیفی، تنظیم و طبقه‌بندی داده‌ها، نمایش ترسیمی، و محاسبه‌ی مقادیری از قییل نما، میانگین، میانه و ... است که حاکی از مشخصات یکایک اعضای جامعه‌ی مورد بحث است. در آمار توصیفی اطلاعات حاصل از یک گروه، همان گروه را توصیف می‌کند و اطلاعات به دست آمده به دسته‌های مشابه تعمیم داده نمی‌شود. به‌طور کلی از سه روش در آمار توصیفی برای خلاصه‌سازی داده‌ها استفاده می‌شود:

۱- استفاده از جدول‌ها

۲- استفاده از نمودارها

۳- محاسبه‌ی مقادیری خاص که نشان‌دهنده‌ی خصوصیات مهمی از داده‌ها باشند.

به بیانی دیگر، آمار توصیفی شامل کلیه‌ی اعمالی است که هدف از آن‌ها خلاصه‌کردن داده‌ها یا توصیف جنبه‌های مهم داده‌هاست، بدون آنکه گامی فراتر از آن برداشته شود که گزارش این خلاصه‌کردن داده‌ها معمولاً به‌صورت عددی یا نموداری انجام می‌گیرد.

چنانچه به جای مطالعه‌ی کل اعضای جامعه، بخشی از آن با استفاده از فنون نمونه‌گیری انتخاب شده، و مورد مطالعه قرار گیرد و بخواهیم نتایج حاصل از آن را به

کل جامعه تعمیم دهیم، از روش‌هایی استفاده می‌شود که موضوع آمار استنباطی است. آن چه که مهم است این است که در گذر از آمار توصیفی به آمار استنباطی یا به عبارت دیگر از نمونه به جامعه، بحث و نقش احتمال شروع می‌شود. در واقع احتمال، پل رابط بین آمار توصیفی و استنباطی به حساب می‌آید. (دائرةالمعارف فارسی)

گرچه آمار توصیفی بخشی مهم از آمار است و هنوز هم استفاده‌ی گسترده‌ای از آن، مخصوصاً برای توصیف ویژگی‌های مهم داده‌ها برای عموم به عمل می‌آید، اما در سال‌های اخیر توجه بیشتر به آمار استنباطی، یعنی روش‌هایی است که به کمک آن‌ها می‌توان اطلاعات موجود در مجموعه‌ای متناهی از داده‌ها را به مجموعه‌ای بزرگ‌تر که داده‌ها از آن به دست آمده‌اند، تعمیم داد.

۱-۳ برخی مفاهیم اولیه

جامعه‌ی (جمعیت) آماری عبارت است از مجموعه‌ای از عناصر مورد نظر که حداقل دارای یک صفت مشخصه‌ی مشترک باشند. مانند جامعه‌ی دانشجویان دانشگاه پیام‌نور، جامعه‌ی ساکنان شهر تهران، جامعه‌ی بیماران قلبی، جامعه‌ی گیاهی و ...

صفت مشخصه صفتی است که بین همه‌ی عناصر آماری مشترک و متمایزکننده‌ی یک جامعه‌ی آماری از سایر جوامع است. مثلاً در جامعه‌ی ایرانیان، ایرانی بودن وجه مشترک و متمایز افراد است.

جامعه‌ی آماری به دو نوع تقسیم می‌شود: متناهی و نامتناهی. اگر مقادیر جامعه از تعداد محدود ثابتی تشکیل شود و پایان‌پذیر باشد، آن را **متناهی** می‌نامیم، مثلاً جامعه‌ی ایران در سرشماری سال ۱۳۹۲ که ۷۷۳۳۶۹۸۰ نفر بوده است، یک جامعه‌ی متناهی است. در مقابل جامعه‌های آماری **نامتناهی** را داریم که تعداد افراد آن بی‌پایان است، مانند تعداد دفعاتی که یک شخص در یک امتحان تا قبولی شرکت می‌کند، که عناصر آن به صورت ۱، ۲، ۳، ... و یک جامعه‌ی آماری نامتناهی است.

از نظر ریاضی، مجموعه‌ای شمارش‌پذیر نامیده می‌شود که با یک زیرمجموعه از اعداد طبیعی در تناظر یک به یک باشد و در غیراین صورت شمارش‌ناپذیر نامیده می‌شود. بنابر آنچه گفته شد برای جامعه‌ی آماری نامتناهی حالت شمارش‌پذیر یا شمارش‌ناپذیر متصور است.

در اکثر مواقع، بررسی کلیه اعضای جامعه، با دو محدودیت وقت و هزینه مواجه است که منجر به نمونه‌گیری از جامعه می‌شود. بخشی از جامعه، که به شیوه‌های آماری انتخاب می‌شود، یک **نمونه** را می‌سازد. مسئله‌ی انتخاب یک نمونه‌ی «خوب» به قدری مهم است که مبحث مهم نمونه‌گیری در تئوری احتمال و آمار به آن اختصاص یافته است. برای مطالعه‌ی ویژگی‌های یک جامعه یا نمونه‌ای از آن باید دانسته‌هایی را که از راه پرسش‌نامه یا طرح آزمایش در آزمایشگاه به دست می‌آیند، به زبان اعداد درآورد و سپس این اعداد را طبق اصولی تجزیه، تحلیل و تفسیر کرد.

صفاتی را که عامل تفاوت‌ها میان افراد جامعه‌ی آماری هستند، **صفات متغیر** می‌نامیم. پس صفات متغیر صفاتی هستند که در یک جامعه از یک عضو به عضو دیگر تغییر می‌کنند. مثلاً در جامعه‌ی دانشجویان دانشگاه پیام‌نور، شهرستان محل سکونت و رشته‌ی انتخابی، صفات متغیر هستند. معمولاً صفاتی نظیر استعداد، هوش، نوع بیماری، گروه خونی، وضع تأهل، سن، طول عمر لامپ، تعداد دفعات تعویض قطعه‌ای از خودرو، میزان تأثیر دارویی خاص و ...، متغیر در نظر گرفته می‌شوند.

انواع متغیر: متغیرها به دو دسته تقسیم می‌شوند: متغیرهای کیفی و متغیرهای کمی.

الف. متغیرهای کیفی: مربوط به صفاتی هستند که واحد نداشته و قابل اندازه‌گیری نیستند مانند جنسیت، مرغوبیت، شغل، نوع بیماری، وضع تأهل، رنگ چشم، گروه خونی و نظایر آنها (همانگونه که در بخش ۱-۴ خواهیم دید، متغیرهای کیفی عمدتاً به گروه متغیرهایی که با مقیاس‌های اسمی و ترتیبی اندازه‌گیری می‌شوند، تعلق دارند).

ب. متغیرهای کمی: مربوط به صفاتی هستند که قابل اندازه‌گیری یا شمارشی هستند؛ یعنی واحد اندازه‌گیری یا شمارش دارند و قابل مقایسه و سنجش هستند. مانند سن، طول عمر، درجات تحصیلی، وزن، قد، درآمد و غیره که از راه اندازه‌گیری به دست می‌آیند. (همانگونه که در بخش ۱-۴ خواهیم دید، متغیرهای کمی عمدتاً به گروه متغیرهایی که با مقیاس‌های فاصله‌ای و نسبتی اندازه‌گیری می‌شوند، تعلق دارند).

متغیرهای کمی خود به دو دسته تقسیم می‌شوند که عبارت‌اند از متغیرهای کمی گسسته و متغیرهای کمی پیوسته. **متغیرهای کمی گسسته** آنهایی هستند که قابل شمارش هستند مانند سن (برحسب تعداد سال‌ها)، تعداد افراد خانوار، ظرفیت مسافر،

تعداد تولید روزانه و جامعه روستاها. متغیرهای کمی پیوسته آنهایی هستند که مقادیرشان در یک بازه از اعداد حقیقی قرار دارد مانند وزن، قد، طول عمر، درآمد، نمره‌ی دروس و هزینه‌ی بیمارستانی برای بیماران.

۱-۴ مفهوم مقیاس‌سازی

منظور از مقیاس‌سازی فرایندی است که به کمک آن به عضوی از جامعه، عدد یا نمره‌ای نسبت داده می‌شود. باید توجه کرد نمره‌هایی (اعدادی) که به این طریق به دست می‌آیند، معانی متفاوتی دارند. برای توضیح متفاوت بودن این نمره‌ها، سه پاسخ ظاهراً یکسان را که به سه پرسش متفاوت داده شده است، در نظر بگیرید. فرض کنید:

کسی که در مسابقات دوی صدمتر کشور شرکت کرده است، به سه پرسش زیر تنها با استفاده از عدد پاسخ می‌گوید:

۱. با چه شماره‌ای در مسابقه شرکت کردید؟

۲. نفر چندم شدید؟

۳. در چند ثانیه به خط پایان رسیدید؟

فرض کنید پاسخ وی به هر سه سؤال «۱۱» باشد. با اینکه در هر سه پاسخ، عدد «۱۱» به کار رفته است، اما این عدد در سه معنی متفاوت به کار برده شده است. اولین ۱۱ به نشانه‌ی شماره‌ی (شناسایی) شرکت‌کننده است، دومی به نشانه‌ی مقام کسب‌شده در مسابقه و سومی به نشانه‌ی «زمان» پایان مسابقه است. بنابراین هر یک از ۱۱ها مجموعه‌ی اطلاعات متفاوتی را شامل می‌شود.

استیونس^۱ (۱۹۴۶ میلادی)، چهار نوع مختلف از مقیاس‌ها را مشخص کرده است که هریک از آنها مجموعه‌ی اطلاعات متفاوتی را در بردارند. در هر مقیاس از عدد استفاده می‌شود ولی اطلاعات حاصل از این اعداد به کلی متفاوت است. این چهار مقیاس عبارت‌اند از: اسمی، ترتیبی، فاصله‌ای و نسبتی. در ادامه به شرح هر یک از آنها می‌پردازیم.

1. Stanley Smith Stevens

۱-۴-۱ مقیاس اسمی

هر مقیاس که تنها برای شناسایی افراد یا اشیاء و یا مکان‌ها به کار می‌رود و معمولاً عددی از مجموعه‌ی اعداد طبیعی است، مقیاس اسمی نامیده می‌شود. برای مثال، گروه‌های خونی افراد عبارت‌اند از: A ، B ، AB ، و O ، که آن‌ها را به ترتیب با اعداد ۱، ۲، ۳ و ۴ مشخص می‌کنیم. این اعداد صرفاً می‌گویند که هر فردی دارای چه گروه خونی است، مورد فردی با برچسب ۳، دارای گروه خونی AB است.

داده‌های اسمی، تنها از نظر ظاهری عددی‌اند، زیرا هیچ‌یک از خاصیت‌های اعدادی را که در حساب معمولی با آنها سروکار داریم از جمله مفهوم بیشتر یا کمتر بودن، را ندارند و اصطلاحاً در مقیاس اسمی، کدگذاری یا اسم‌گذاری انجام می‌شوند و اعداد نسبت داده شده فاقد خاصیت ترتیب هستند. در روش‌های تحلیلی آماری توجه به اینکه آیا به کار بردن اعمال حسابی در مورد داده‌ها مجاز است یا خیر، اهمیت زیادی دارد.

۱-۴-۲ مقیاس ترتیبی

هرگاه مقیاس، یک عدد حقیقی باشد که صرفاً برتری را بیان کند، آن را مقیاس ترتیبی می‌گویند. مواردی پیش می‌آید که صرف‌نظر از تفاوت محتویات یک طبقه با محتویات طبقه‌ی دیگر، یک نوع ارتباط ترتیبی بین آنها برقرار است. روابط موجود بین طبقات با توجه به نوع مقیاس، اکثر مواقع به صورت "بالتر، بیشتر، مشکل‌تر و" بیان می‌شود که چنین روابطی را معمولاً با علائمی نظیر " $>$ " نشان می‌دهند. استفاده از اصطلاحاتی چون "بالا، وسط، پایین" و "قوی، متوسط، ضعیف" در تحقیقات بیانگر مقیاس ترتیبی است. در این نوع مقیاس، مانند مقیاس اسمی، مجاز به انجام عملیات جبری نیستیم.

در مقیاس ترتیبی وقتی اعداد ۲، ۴، ۶، ۸، ۱۰ را به ترتیب به پنج فرد A ، B ، C ، D ، E اختصاص می‌دهیم به معنی آن است که پنج فرد مذکور از لحاظ صفت مورد نظر به ترتیب A ، B ، C ، D ، E قرار می‌گیرند. لذا اگر وضع نسبی افراد یا اشیاء را در مقیاس ترتیبی بدانیم، می‌توانیم آن‌ها را به صورت‌های زیر نمایش دهیم: $B > A$ ؛ یعنی، B از A بزرگتر است. $B < C$ ؛ یعنی، B از C کوچکتر است. $B = C$ ؛ یعنی، B با C هم‌ارزش است. برای مثال، در یک مسابقه‌ی دو میدانی دوندگاری را که به ترتیب اول، دوم و سوم به خط پایان می‌رسند، با اعداد ۱، ۲ و ۳ مشخص کنیم. در

این حالت اعداد نسبت داده شده دارای خاصیت ترتیب بوده ولی فاقد خاصیت نسبت هستند.

۱-۴-۳ مقیاس فاصله‌ای

وقتی که مقیاسی همه‌ی خصوصیات یک مقیاس ترتیبی را دارا باشد و علاوه بر آن فاصله‌ی بین هر دو عدد بر روی مقیاس، میزان مشخصی داشته باشد، همچنین نسبت تفاضل با دو فاصله را حفظ کند، آن را مقیاس فاصله‌ای می‌نامند. در این نوع اندازه‌گیری نسبت هر دو فاصله مستقل از واحد اندازه‌گیری و مستقل از نقطه‌ی صفر است. در مقیاس فاصله‌ای مکان نقطه‌ی صفر و واحد اندازه‌گیری اختیاری و قراردادی است.

در اندازه‌گیری دما، با وجود آن‌که دما در دو مقیاس سانتی‌گراد و فارنهایت دارای نقطه‌ی صفر متفاوت است، ولی هر دو برای اندازه‌گیری حرارت به‌کار می‌روند. برای مثال، پنج کمیت قابل اندازه‌گیری شده X_1, X_2, X_3, X_4, X_5 را در نظر می‌گیریم، این کمیت‌ها را به اندازه‌های ۱۱، ۱۴، ۱۷، ۲۰، ۲۳ مقیاس‌سازی کرده‌ایم (یعنی X_i ها در عددی ثابت ضرب و با عددی ثابت جمع شده‌اند و داده‌های جدید به دست آمده است). ملاحظه می‌شود که اختلاف اندازه‌ی کمیت دوم از اول $(3 = 11 - 14)$ برابر سه برابر اختلاف کمیت اول از چهارم $(9 = 11 - 20)$ است.

مثال دیگری از مقیاس فاصله‌ای تبدیل تقویم سال هجری قمری به تقویم سال هجری خورشیدی است که سال قمری را در ۱۱ ضرب کرده، حاصل را بر ۳۶۵ تقسیم می‌کنند. سپس نتیجه را از همان سال قمری تفریق می‌کنند. تبدیل به سایر گاه‌شمارها نیز از این نوع مقیاس محسوب می‌شود.

۱-۴-۴ مقیاس نسبتی

به مقیاسی که علاوه بر همه‌ی خصوصیات مقیاس فاصله‌ای، دارای نقطه‌ی صفر واقعی نیز هست، مقیاس نسبتی می‌گویند. در مقیاس نسبتی، نسبت هر دو قطعه روی مقیاس اندازه‌گیری مستقل از واحد اندازه‌گیری است.

مقیاس‌هایی چون پوند و گرم، نقطه‌ی صفر واقعی دارند و نسبت هر دو وزن دل‌خواه از واحد اندازه‌گیری مستقل است. بیشتر اطلاعاتی که با استفاده از ابزارهای

فیزیکی اندازه‌گیری می‌شود، از نوع مقیاس نسبتی است. مثلاً، اگر در سطح مقیاس نسبتی، اعداد ۲، ۴، ۶، ۸، ۱۰ نماینده‌ی مقدار صفت معینی در اشیای A ، B ، C ، D ، E باشند، می‌توان گفت که این صفت در B به اندازه‌ی ۲ برابر A و در C به اندازه‌ی ۳ برابر A و در E ، ۵/۲ برابر B وجود دارد.

جدول ۱-۱ نشان‌دهنده‌ی ویژگی‌های ذکر شده درباره‌ی مقیاس‌های چهارگانه و مقایسه‌ی آنها است.

جدول ۱-۱. مقایسه‌ی چهار نوع مقیاس اندازه‌گیری

مقیاس					
نسبتی	فاصله‌ای	ترتیبی	اسمی		
دارد	دارد	دارد	ندارد	ترتیب	ویژگی
دارد	دارد	ندارد	ندارد	فواصل	
دارد	دارد	ندارد	ندارد	مبدأ صفر قراردادی	
دارد	ندارد	ندارد	ندارد	مبدأ صفر مطلق	

۱-۵ انواع داده‌ها

یک مجموعه‌ی عددی را که با استفاده از مقیاس مناسب اندازه‌گیری صفتی به‌دست آمده است، داده می‌نامیم. نوع داده با نوع متغیر مورد مطالعه متناظر است؛ پس از تقسیم‌بندی انواع متغیر در بخش ۱-۳ به دو نوع گسسته و پیوسته، در اینجا نیز داده‌ها به دو نوع گسسته و پیوسته تقسیم می‌شوند. داده‌های گسسته از راه اندازه‌گیری با مقیاس‌های اسمی، ترتیبی یا شمارشی به‌دست می‌آیند. داده‌های گسسته را داده‌های «جدا» نیز می‌نامند. داده‌های پیوسته از راه اندازه‌گیری با مقیاس‌های فاصله‌ای یا نسبتی به‌دست می‌آیند. عمومی‌ترین داده‌ی پیوسته به‌وسیله‌ی ابزارهای اندازه‌گیری برحسب سانتی‌متر (برای فاصله)، کیلوگرم (برای سنگینی)، ثانیه (برای زمان)، فارنهایت (برای دما) و یا واحدهای تبدیل یافته‌ی آنها حاصل شده است.

معمولاً داده‌های حاصل از نمونه‌گیری به اندازه‌ی n را با $x_1, x_2, \dots, x_n$ و واحدهای حاصل از سرشماری به اندازه‌ی N را با $x_1, x_2, \dots, x_N$ نشان می‌دهیم.

با توجه به گستردگی مباحث علمی، محققان منابع گوناگونی برای گردآوری داده‌ها در اختیار دارند. بدون وارد شدن در جزئیات به انواع مهم داده‌ها برحسب منابع آنها اشاره می‌کنیم.

داده‌های اولیه اطلاعات خامی هستند که به‌طور معمول از طریق **سنجش مستقیم**، مانند انجام مصاحبه، مشاهدات میدانی، اندازه‌گیری با ابزارهای ویژه، به‌دست می‌آیند، یا از طریق **سنجش غیرمستقیم** که در آن سنجش و ثبت نتایج در مکان‌های مجزا از هم انجام می‌شود و این روش استفاده از مصاحبه‌های تلفنی، شمارنده‌های ترافیک، و مشخصه‌های بارندگی خودکار را شامل می‌شود. داده‌های حاصل از روش‌های سنجش از راه دور مانند دریافت اطلاعات از ماهواره‌ها نیز در زمره‌ی داده‌های اولیه محسوب می‌شوند.

داده‌های ثانویه که در آن اطلاعات که پیشتر- به هر منظور خاص- گردآوری شده‌اند، منبع اصلی تلقی شده و اطلاعات مربوط به تحلیل موردنظر از آن استخراج می‌شود. این داده‌ها یا پیامدهای غیرمستقیم فعالیت‌های دستگاه‌های اجرایی دولتی یا خصوصی، نظیر ادارات ثبت احوال و ثبت اسناد، مؤسسات نقشه‌برداری، بانک‌ها، شهرداری‌ها، و غیره‌اند یا داده‌های حاصل از سرشماری‌های رسمی سراسری و یا مقطعی مؤسساتی نظیر مرکز آمار ایران. داده‌های مندرج در تصاویر حاصل از دستگاه‌های سنجش از راه دور نیز از این جمله‌اند. ماهواره‌ی طلوع که دارای فناوری‌هایی مانند محموله‌ی تصویربرداری، کنترل وضعیت مکانیزم‌ها و سلول‌های خورشیدی است و مأموریت اصلی آن، تصویربرداری تک‌طیفی، ذخیره و ارسال داده‌های تصویر به ایستگاه‌های زمینی می‌باشد، یک نمونه‌ی سنجش از راه دور است.

بدیهی‌است که تفاوت‌های عمده‌ای بین فرایندهای اندازه‌گیری مربوط به منابع گوناگون، و بین داده‌های اولیه و ثانویه وجود دارد. وقتی از داده‌های اولیه استفاده می‌کنیم، معمولاً طرح پژوهشی مربوط به تحلیل آسان‌تر است اما فرایند اندازه‌گیری معمولاً دشوار و وقت‌گیر است. با توجه به امکان تحت کنترل داشتن فرایند گردآوری داده‌ها در این حالت، می‌توان متغیرها را به هر شکلی که در تحلیل آماری مورد نیاز باشند، انتخاب و اندازه‌گیری کرد. اما، وقتی از داده‌های ثانویه استفاده می‌کنیم، کنترلی نسبت به متغیرها و فرایند اندازه‌گیری نداریم و این امر ممکن است موجب ایجاد

مشکلاتی در موقع تجزیه و تحلیل شود. در این صورت لازم است محقق ابتدا از مناسب بودن داده‌ها اطمینان حاصل کند.

خلاصه‌ی فصل اول

در آغاز این فصل با بیان برخی رویدادهای تاریخی و پیشینه اهمیت دانش آماری اشاره شد به طوری که با گسترش و پدیدار شدن کاربردهای متعدد، شاخه‌ای از علم ریاضی شکل گرفت که آمار نامیده می‌شود. علم آمار به دو بخش آمار توصیفی و آمار استنباطی تقسیم شده است. آمار توصیفی، تنظیم و طبقه‌بندی داده‌ها، نمایش ترسیمی، و محاسبه‌ی مقادیری از قبیل نما، میانگین، میانه و ... است. برای انجام آمار توصیفی لازم است ابتدا بر روی مشاهدات مقیاس‌سازی را انجام دهیم. نوع مقیاس (اسمی، ترتیبی، فاصله‌ای، نسبی) در محاسبه‌ی کمیت‌هایی که در آمار توصیفی انجام می‌شوند اهمیت دارد. آمار استنباطی روش‌هایی را در برمی‌گیرد که برپایه‌ی علم احتمالات، با استدلال استقرایی، نتایج را از جزء (نمونه) به کل (جامعه) تعمیم دهد.

تمرین

۱-۱. مدیریت ماشین‌سازی تبریز قصد دارد میانگین عمر قطعه‌ای خاص تولیدی که توسط مشتریان خریداری شده است را برآورد کند. برای این بررسی، جامعه و نمونه را تعریف کنید. متغیر مورد مطالعه چه نوع متغیری است؟

۱-۲. محقق می‌ایل است دو نوع روش درمانی گیاهی و صنعتی را بر روی بیماران مبتلا به فشار خون بالا در شهر تهران بررسی کند. او از بین بیماران داوطلب، نمونه‌ای تصادفی به اندازه‌ی ۳۰ نفر را انتخاب کرده و به‌طور تصادفی آنها را در دو گروه ۱۵ نفری تحت درمان دوماهه قرار می‌دهد. جامعه و متغیر مورد بررسی در این تحقیق کدام است؟

۱-۳. وزارت آموزش و پرورش می‌خواهد مقایسه‌ای از نظر میزان یادگیری مطالب درسی بین دو نظام آموزشی موجود و قدیمی را انجام دهد. جامعه‌ی آماری و نمونه‌ی این تحقیق را مشخص کنید.

۱-۴. ویژگی‌های زیر را با چه مقیاسی می‌توان اندازه‌گیری کرد؟ کدام یک از متغیرها گسسته و کدام پیوسته‌اند؟

دست مزد	میزان سواد	ارتفاع از سطح دریا
زمان	وضعیت تاهل	طول جغرافیایی
جنسیت	ارتفاع	وزن ماهی
رضایت از شغل	عمق	طول بال پرنده
رضایت از درآمد	عرض جغرافیایی	درجه حرارت

خودآزمایی چهارگزینه‌ای

۱- کدام گزینه یک متغیر پیوسته است؟

الف. مدرک تحصیلی ب. وضعیت تأهل ج. سن د. جنسیت

۲- کدام گزینه یک متغیر گسسته است؟

الف. تعداد درختان یک شهر ج. میزان تأثیر یک دارو در درمان نوعی بیماری

ب. طول عمر یک لامپ د. میزان کاهش آب در دریاچه‌ای معین

۳- اگر x مقدار یک متغیر باشد، آنگاه تبدیل $y = ax, a > 0$ در چه مقیاسی به کار

می‌رود؟

الف. اسمی ب. ترتیبی ج. نسبی د. فاصله‌ای

۴- اگر x مقدار یک متغیر باشد، آنگاه تبدیل $y = ax + b, a > 0, b \neq 0$ در چه

مقیاسی به کار می‌رود؟

الف. اسمی ب. ترتیبی ج. نسبی د. فاصله‌ای

۵- کدام ویژگی از مقیاس اسمی است؟

الف. مصرف سوخت ب. محبوبیت ج. رنگ د. میزان یادگیری حیوانات

۶- وزن محصولات تولیدی از یک شرکت، دارای چه نوع مقیاسی است؟

الف. رتبه‌ای ب. فاصله‌ای ج. اسمی د. نسبی

۷- آمار توصیفی

الف. محاسبه‌ی معیارها از روی نمونه است و به توصیف آنها می‌پردازد.

ب. از روی معیارهای به‌دست آمده نتایج را به جامعه تعمیم می‌دهد.

ج. توصیفی از وضعیت موجود نمونه است.

د. اعتبار نمونه‌ی تصادفی را می‌سنجد.

۱- ۸. کدام یک از گزینه‌های زیر صفت کمی پیوسته هستند؟

- الف. گروه خونی
ب. تعداد افراد خانواده
ج. سن افراد به سال
د. ظرفیت مسافران یک خودرو

۱- ۹. کدام یک از صفات زیر کیفی است؟

- الف. مدرک تحصیلی
ب. عنوان شغلی
ج. جنسیت
د. تمام متغیرها

۱- ۱۰. کدام یک از صفات زیر کیفی است؟

- الف. دستمزد
ب. هزینه بیمارستان
ج. وضع تأهل
د. رتبه‌ی شرکت‌کنندگان در یک مسابقه

فصل دوم

جدول و نمودارهای آماری

هدف کلی

آموزش، براساس مشاهدات، ساخت جدول آماری و رسم نمودارهای آماری مناسب جهت ارایه‌ی توصیف دقیق و منطقی برای جامعه‌ی آماری

هدف‌های یادگیری

- خواننده پس از یادگیری این فصل باید بتواند:
- اهمیت نمودارها و جداول آماری در توصیف جامعه‌های آماری را بیان کند.
- مرتب‌سازی داده‌ها را انجام دهد.
- براساس ارقام مشاهدات، نمودار ساقه و برگ را رسم کند.
- تعداد رده‌ها در یک جدول آماری را تعیین کند.
- نماینده‌ی هر رده را تعیین کند.
- از روی ستون فراوانی، ستون‌های فراوانی نسبی، فراوانی تجمعی و فراوانی تجمعی نسبی را محاسبه کند.
- براساس مشاهدات از یک متغیر پیوسته، نمودار بافت‌نگار را ترسیم کند.
- نمودار چندبر فراوانی و اوجایو را رسم نماید.
- نمودار میله‌ای، برای مشاهدات از روی یک متغیر گسسته را رسم کند.
- سهم قطاع هر رده در نمودار دایره‌ای را محاسبه و نمودار را رسم کند.

مقدمه

نمودارهای آماری فرایند کشف در تحلیل داده‌ها با آشکارسازی موارد غیرقابل انتظار و یافتن ساختارهای پنهان را تقویت می‌کنند. مبحث نمودارها اغلب مرتبط با تحلیل اکتشافی داده‌ها یا کشفیات در خلال داده‌کاوی است که امروزه در دنیای تحلیل داده‌ها فعالیت مهم و ارزشمندی به شمار می‌رود. توکی^۱ (۱۹۷۷) در این زمینه پیشگام بود. نمودارها در مواردی از قبیل درستی مدل، تشخیص روابط و ارتباط بین یافته‌ها که اطلاعات عددی کافی نیستند، مورد استفاده قرار می‌گیرند. در سال‌های اخیر پیشرفت‌های جدیدی در تحقیق نموداری حاصل شده است. مطالعات اولیه‌ی ارزیابی اینکه چگونه نمودارهای آماری توسط چشم انسان خوانده و درک می‌شوند را در کتابنامه بازرگان‌لاری، ۱۳۷۵، جانسون^۲ و کاتز^۳، ۱۹۷۵ می‌توان یافت.

برای آنکه بتوان نخستین گام را برای تجزیه و تحلیل داده‌ها برداشت و نتایجی درباره‌ی خصوصیات جامعه‌ای که این اطلاعات از آن استخراج شده‌اند، به دست آورد، لازم است که این داده‌ها سازماندهی (تنظیم) و تلخیص شوند. در این فصل، دو بخش مهم از آمار توصیفی که به ساخت جداول آماری و نمودارهای آماری اختصاص دارد را شرح می‌دهیم. آنچه از دیدگاه مباحث آمار توصیفی مدنظر است دو خلاصه گزارش اطلاعات یکی در قالب جدول، و مرحله‌ی بعدی با نمایش یک شکل هندسی بیان می‌شود.

اولین گام، مرتب ساختن اطلاعات است تا بتوان آنها را در جداول آماری خلاصه کرد. مراحل تعریف رده‌های یک جدول آماری و سپس تخصیص هریک از مشاهدات به رده‌ای خاص را طی محاسباتی بسیار ساده در بخش ۲-۱ انجام خواهیم داد.

اخیراً نمودارهای آماری نقش مهمی را در تحلیل اکتشافی داده‌ها ایفا می‌کنند که نتایج تحقیقات را می‌توانید در کیلیولند^۴ و مک‌گیل^۵، ۱۹۸۵ ملاحظه کنید. همچنین سعی می‌شود آن دسته از نمودارهای آماری را در متن این فصل بیاوریم که به بهترین وجه، کار توصیف اطلاعات را انجام دهد. خواننده چنانچه به مهارت‌های رایانه‌ای تسلط داشته باشد (مثلاً مسلط به محیط‌های آفیس)، به‌خوبی می‌تواند در قسمت «درج

1. Tukey 2. Johnson 3. Kotz 4. Cleveland 5. McGill

چارت^۱» با نمودارهای متنوع دیگری نیز آشنا شود. لازم به ذکر است در برخی متون مانند هیر^۲ و بوستوک^۳، ۲۰۱۰ و ماجودر^۴ و همکاران، ۲۰۱۳ آموزش نمودارهای دیگری، علاوه بر آنچه که در این کتاب آمده است، را می توان مورد مطالعه قرار داد.

۲-۱ مرتب کردن داده های عددی

ابتدا به مثال زیر توجه کنید.

مثال ۲-۱. فرض کنید داده های زیر میزان بارندگی سالانه ی ۱۰۰ شهر را برحسب سانتی متر در منطقه ای نشان بدهد.

جدول ۲-۱. میزان بارندگی سالانه ۱۰۰ شهر

۴۴٫۰	۳۵٫۴	۲۸٫۴	۳۷٫۰	۴۶٫۰	۳۵٫۴	۱۹٫۴	۲۰٫۴	۵۶٫۴	۴۳٫۲
۳۶٫۲	۳۸٫۴	۴۹٫۲	۳۱٫۸	۸۶٫۴	۱۲٫۶	۲۷٫۴	۱۴٫۰	۳۹٫۴	۳۹٫۴
۱۵٫۸	۲۸٫۸	۳۸٫۰	۴۴٫۰	۳۸٫۴	۷۴٫۰	۲۳٫۰	۱۱٫۴	۳۹٫۸	۳۰٫۲
۲۹٫۲	۴۰٫۶	۴۹٫۶	۳۰٫۴	۱۲٫۲	۱۲۳٫۸	۴۲٫۰	۴۷٫۰	۳۲٫۴	۳۹٫۲
۳۵٫۲	۵۶٫۴	۳۱٫۰	۴۵٫۰	۹۰٫۲	۱۰۰٫۰	۳۹٫۰	۳۷٫۰	۴۹٫۴	۲۸٫۲
۱۲٫۶	۲۷٫۰	۴۷٫۸	۵۲٫۶	۴۱٫۰	۴۰٫۰	۲۸٫۰	۲۳٫۶	۳۷٫۶	۳۷٫۸
۳۰٫۰	۴۸٫۵	۱۸٫۰	۴۱٫۰	۲۲٫۶	۲۴٫۲	۸۹٫۶	۹۰٫۴	۴۳٫۰	۲۹٫۸
۵۶٫۲	۲۴٫۸	۱۲٫۶	۵۳٫۶	۱۲۵٫۴	۱۶٫۲	۳۹٫۰	۴۰٫۸	۳۳٫۶	۳۹٫۴
۴۵٫۶	۳۷٫۴	۱۸٫۰	۵۰٫۶	۱۰۳٫۴	۵۲٫۴	۲۰٫۲	۶۴٫۶	۲۲٫۲	۶۰٫۰
۴۲٫۲	۴۲٫۰	۱۶٫۲	۱۰۸٫۲	۴۴٫۰	۴۲٫۶	۳۹٫۴	۳۷٫۶	۴۱٫۴	۴۰٫۴

شاید در نخستین مرحله، بهتر این باشد که مقادیر فرین، یعنی بیشترین و کمترین مقدار داده ها را در این مجموعه معین کنیم. با یک بررسی ساده معلوم می شود که کمترین مقدار ۱۱٫۴ و بیشترین مقدار ۱۲۵٫۴ است. می توان همین شیوه را ادامه داد و دومین مقدار، سومین مقدار، و ... را از نظر بزرگی پیدا کرد یا به عبارت دیگر داده ها را به ترتیب صعودی مرتب کرد (می توان داده ها را به ترتیب نزولی نیز مرتب کرد). پس از انجام این کار فهرست زیر را خواهیم داشت:

با اینکه مرتب کردن دستی داده ها به ترتیب نزولی یا صعودی، برای مجموعه های بزرگ کار چندان آسانی نیست، ملاحظه می کنیم که این مجموعه ی جدید، چندان

تفاوتی با داده‌های قبلی ندارد و نمی‌توان تنها با نظر کردن به آنها از کم و کیف «توزیع» آنها اطلاعی به دست آورد. بنابراین سودمند خواهد بود که راهی دیگر برای پردازش داده‌ها پیدا کنیم. یکی از روش‌هایی که اخیراً به وجود آمده است، نمودار ساقه و برگ است که تصویری از وضع کلی داده‌ها را در اختیار می‌گذارد. رسم نمودار ساقه و برگ را در مثال ۲-۲ توضیح می‌دهیم.

جدول ۲-۲. مرتب شده‌ی داده‌های میزان بارندگی سالانه ۱۰۰ شهر

۱۱٫۴	۱۸٫۰	۲۷٫۰	۳۰٫۴	۳۷٫۰	۳۹٫۲	۴۱٫۰	۴۴٫۰	۴۹٫۶	۷۴٫۰
۱۲٫۲	۱۹٫۴	۲۷٫۴	۳۱٫۰	۳۷٫۴	۳۹٫۴	۴۱٫۰	۴۴٫۰	۵۰٫۶	۸۶٫۴
۱۲٫۶	۲۰٫۲	۲۸٫۰	۳۱٫۸	۳۷٫۶	۳۹٫۴	۴۱٫۴	۴۵٫۰	۵۲٫۴	۸۹٫۶
۱۲٫۶	۲۰٫۴	۲۸٫۲	۳۲٫۴	۳۷٫۶	۳۹٫۴	۴۲٫۰	۴۵٫۶	۵۲٫۶	۹۰٫۲
۱۲٫۶	۲۲٫۲	۲۸٫۴	۳۳٫۶	۳۷٫۸	۳۹٫۴	۴۲٫۰	۴۶٫۰	۵۳٫۶	۹۰٫۴
۱۴٫۰	۲۲٫۶	۲۸٫۸	۳۵٫۲	۳۸٫۰	۳۹٫۸	۴۲٫۲	۴۷٫۰	۵۶٫۲	۱۰۰٫۰
۱۵٫۸	۲۳٫۰	۲۹٫۲	۳۵٫۴	۳۸٫۴	۴۰٫۰	۴۲٫۶	۴۷٫۸	۵۶٫۴	۱۰۳٫۴
۱۶٫۲	۲۳٫۶	۲۹٫۸	۳۵٫۴	۳۸٫۴	۴۰٫۴	۴۳٫۰	۴۸٫۵	۵۶٫۴	۱۰۸٫۲
۱۶٫۲	۲۴٫۲	۳۰٫۰	۳۶٫۲	۳۹٫۰	۴۰٫۶	۴۳٫۲	۴۹٫۲	۶۰٫۰	۱۲۳٫۸
۱۸٫۰	۲۴٫۸	۳۰٫۲	۳۷٫۰	۳۹٫۰	۴۰٫۸	۴۴٫۰	۴۹٫۴	۶۴٫۶	۱۲۵٫۴

مثال ۲-۲. فرض کنید داده‌های زیر مربوط به تعداد شعبه‌های یک بانک در بیست شهر باشند.

جدول ۳-۲. تعداد شعبه‌های یک بانک در بیست شهر

۶۹	۸۴	۵۲	۹۳	۶۱	۷۴	۷۹	۶۵	۸۸	۶۳
۵۷	۶۴	۶۷	۷۲	۷۴	۵۵	۸۲	۶۱	۶۸	۷۷

پس از مرتب کردن داده‌ها، رقم یکان و دهگان هر عدد را جداگانه در نظر بگیرید و اعدادی را که رقم دهگان آنها یکی هستند، یک جا گردآوری کنید. برای این منظور مثلاً عدد ۶۹ را به صورت ۶|۹ در نظر می‌گیریم. رقم دهگان، یعنی ۶ را در ستون اول در جای مناسب خود می‌نویسیم و رقم یکان را در سمت راست علامت | وارد می‌کنیم. برای نوشتن ۶۴، چون قبلاً ۶ را در ستون اول نوشته‌ایم، کافی است تنها ۴، یعنی رقم

جدول و نمودارهای آماری ۲۱

یکان را در سمت راست علامت | وارد کنیم. به این ترتیب اعداد به شکل زیر گروه‌بندی می‌شوند:

۵	۲	۷	۵						
۶	۹	۱	۵	۳	۴	۷	۱	۸	
۷	۲	۹	۴	۴	۷				
۸	۴	۸	۲						
۹	۳								

اولین سطر این جدول، یعنی ۵ | ۲ ۷ ۵ نشان می‌دهد که این فهرست شامل اعداد ۵۲، ۵۷ و ۵۵ است و دومین سطر بیان می‌کند که فهرست شامل ۸ عدد است که رقم دهگان آنها ۶ است.

این جدول، نمایش ساقه و برگ نامیده می‌شود، زیرا هر سطر حالت ساقه‌ای را نشان می‌دهد، و هر رقم واقع در سمت راست خط عمودی | حالت یک برگ را نشان می‌دهد. برای تشکیل دادن این جدول، ابتدا با ساقه‌ها، یعنی:

۵
۶
۷
۸
۹

شروع می‌کنیم. سپس با مراجعه به فهرست داده‌ها، به ترتیب هر عدد را در نظر گرفته و دومین رقم آن را به عنوان برگ در روی ساقه‌ی مربوط به آن درج می‌کنیم. نمایش ساقه و برگ همان اطلاعاتی را منعکس می‌سازد که در فهرست اصلی موجود است اما بسیار فشرده‌تر از آن است. بخش عمده‌ی کار مرتب‌کردن داده‌ها، با ایجاد نمایش ساقه و برگ انجام می‌شود. با مرتب‌کردن برگ‌ها که کار نسبتاً ساده‌تری است، می‌توان کار مرتب‌کردن داده‌ها را گامی جلوتر برد. در این مثال جدول زیر را به دست می‌آوریم:

۵	۲	۵	۷						
۶	۱	۱	۳	۴	۵	۷	۸	۹	
۷	۲	۴	۴	۷	۹				
۸	۲	۴	۸						
۹	۳								

در واقع، نمایش ساقه و برگ و سپس مرتب کردن برگ‌ها شبیه کاری است که در فرهنگ‌های لغت انجام می‌شود، واژه‌ها ابتدا برحسب اولین حرف خود مرتب می‌شوند و سپس برحسب حروف دوم و سوم و ... البته در این حالت مانند این است که کلمه دوحرفی است.

توجه می‌کنیم که قرارگرفتن هر فهرستی از داده‌ها در یک نمایش ساقه و برگ به سادگی آنچه در مثال اخیر انجام شد، نیست. گاهی با زیادشدن ارقام و پراکندگی داده‌ها می‌توان از گردشده‌ی داده‌ها یا حذف برخی رقم‌های اعشاری، استفاده کرد. راهنمای رسم نمودار ساقه و برگ مثال ۲-۱، با استفاده از نرم‌افزار «آر» در فصل هفتم آمده است.

۲-۲ توزیع‌های فراوانی

تعداد داده‌هایی که در یک مطالعه‌ی آماری گردآوری می‌شوند، معمولاً بیشتر از آن است که تنها با نگاه کردن به آنها، بتوان اطلاعی درباره‌ی پدیده‌ی مورد مطالعه به دست آورد و لازم است که این داده‌ها را به نحوی سازمان دهیم. یک راه ساده برای این منظور، گردآوردن آن‌ها در قالب یک نمایش ساقه و برگ است. اما برای مطالعه دقیق‌تر و مفصل‌تر داده‌ها، لازم است آن‌ها را با دقت بیشتری سازماندهی کنیم. یکی از روش‌های بسیار رایج، استفاده از **توزیع‌های فراوانی** برای رده‌بندی یا گروه‌بندی داده‌هاست. به عنوان مثال فرض کنید که در یک نمونه‌گیری از دانشجویان دانشگاه پیام‌نور در مرکز شیراز، ۴۱ نفر رشته‌ی مهندسی کامپیوتر، ۴۷ نفر رشته‌ی علوم کامپیوتر، ۲۸ نفر رشته‌ی ریاضی کاربردی، و ۱۱ نفر رشته‌ی ریاضی محض بوده‌اند. دسته‌بندی این افراد به صورت جدول ۲-۴ قابل نمایش است.

جدول ۲-۴. یک مثال از توزیع رشته‌ای

عنوان رشته	تعداد
مهندسی کامپیوتر	۴۱
علوم کامپیوتر	۴۷
ریاضی کاربردی	۲۸
ریاضی محض	۱۱

جدول و نمودارهای آماری ۲۳

چنین جدولی، یک توزیع رسته‌ای یا توزیع کیفی نامیده می‌شود و برای دسته‌بندی داده‌های غیرعددی در قالب رسته‌ها از آن استفاده می‌شود. معمولاً برای سازمان‌دادن داده‌های غیرعددی یا داده‌هایی که برحسب مقیاس اسمی بیان شده‌اند، از توزیع‌های رسته‌ای استفاده می‌کنیم.

برای داده‌های عددی برحسب مقیاس‌های دیگر، از توزیع‌های کمی یا توزیع‌های عددی استفاده می‌کنیم. در این حالت، به طوری که خواهیم دید، داده‌ها را در بازه‌ها یا فاصله‌هایی از اعداد قرار می‌دهیم. اعم از اینکه برای داده‌ها توزیع‌های رسته‌ای یا عددی تشکیل داده باشیم یا خیر، خواهیم گفت که داده‌ها را در یک جدول فراوانی خلاصه کرده‌ایم. در مثال ۲-۳ چگونگی تشکیل جدول‌های فراوانی را برای داده‌های کمی-پیوسته به تفصیل بررسی می‌کنیم.

مثال ۲-۳. داده‌های جدول ۲-۵ میزان اکسید سولفور منتشر شده در هوا در ۸۰ روز (برحسب تن) را نشان می‌دهند.

جدول ۲-۵. داده‌های میزان اکسید سولفور منتشره در هوا (برحسب تن) در ۸۰ روز

۱۵/۸	۲۶/۴	۱۷/۳	۱۱/۲	۲۳/۹	۲۴/۸	۱۸/۷	۱۳/۹	۹/۰	۱۳/۲
۲۲/۷	۹/۸	۶/۲	۱۴/۷	۱۷/۵	۲۶/۱	۱۲/۸	۲۸/۶	۱۷/۶	۲۳/۷
۲۶/۸	۲۲/۷	۱۸/۰	۲۰/۵	۱۱/۰	۲۰/۹	۱۵/۵	۱۹/۴	۱۶/۷	۱۰/۷
۱۹/۱	۱۵/۲	۲۲/۹	۲۶/۶	۲۰/۴	۲۱/۴	۱۹/۲	۲۵/۶	۱۶/۹	۱۹/۰
۱۸/۵	۲۳/۰	۲۴/۶	۲۰/۱	۱۶/۲	۱۸/۰	۷/۷	۱۳/۵	۲۳/۵	۱۴/۵
۱۴/۴	۲۹/۰	۱۹/۴	۱۷/۱	۲۰/۸	۲۴/۳	۲۲/۵	۲۴/۶	۱۸/۴	۱۸/۱
۸/۳	۲۱/۹	۱۲/۳	۲۲/۳	۱۳/۳	۱۱/۸	۱۹/۳	۲۰/۰	۲۴/۷	۳۱/۸
۲۵/۹	۱۰/۵	۱۵/۹	۲۷/۵	۱۸/۱	۱۷/۹	۹/۴	۲۴/۱	۲۰/۱	۲۸/۵

برای تشکیل جدول فراوانی (پس از مرتب‌سازی داده‌ها) در نخستین گام باید تعداد رده‌ها یا گروه‌هایی را تعیین کنیم که می‌خواهیم داده‌ها را در آنها جای دهیم. تعیین تعداد رده‌ها تا حدی به سلیقه و نظر محقق بستگی دارد. اما چندین قاعده‌ی کلی برای

انتخاب تعداد رده‌ها پیشنهاد شده است که از آن جمله می‌توان استفاده از هریک از فرمول‌های

$$n = 2^k \quad (۲-۱)$$

$$k = 1 + 3/2 \log(n) \quad (۲-۲)$$

را ذکر کرد که در آن n تعداد داده‌ها و k تعداد رده‌هاست. فرمول (۲-۲) به «دستور استورج» معروف است. آنچه همواره باید به‌خاطر داشت این است که:

الف. به ندرت تعداد رده‌ها را کمتر از پنج و بیشتر از پانزده انتخاب می‌کنیم،

ب. تعداد دقیق رده‌ها به تعداد اندازه‌گیری‌ها یا مشاهدات بستگی دارد.

وقتی تعداد رده‌ها را بزرگ انتخاب می‌کنیم، با همان مشکلی روبه‌رو هستیم که قبل از تلخیص کردن داده‌ها و کارکردن با داده‌های اصلی در پیش است. وقتی تعداد رده‌ها کم باشد، اطلاعات نسبتاً قابل‌توجهی درهم ادغام و از دقت نتیجه‌گیری کاسته می‌شود. برای داده‌های مثال حاضر، چون $n = 80$ ، مطابق فرمول (۲-۲)،

$$k = 1 + 3/2 \log(80) = 7.28$$

لذا می‌توان تعداد رده‌ها را ۷ گرفت.

پس از تعیین تعداد رده‌ها، باید خود رده‌ها یا طبقه‌هایی که داده‌های ما را در برگیرند، مشخص کنیم. در تعیین رده‌ها به قواعد زیر توجه می‌کنیم:

۱. رده‌ها باید طوری انتخاب شوند که همه‌ی داده‌ها را دربرگیرند.

۲. حتماً باید مطمئن شد که هر داده تنها در یک رده قرار می‌گیرد، به عبارت دیگر رده‌ها نباید عضو مشترک داشته باشند.

۳. در صورت امکان طول رده‌ها را برابر انتخاب می‌کنیم، و به علاوه سعی می‌کنیم که **کران رده‌ها یا حدود رده‌ها** مضربی از ۵، ۱۰، ۱۰۰، ... باشند. منظور از حدود هر رده، کمترین و بیشترین مقدار از داده‌هاست که می‌توانیم در رده‌ای معین قراردهیم.

در این مثال، چون کمترین مقدار داده‌ها $\min = 6.2$ و بیشترین مقدار داده‌ها $\max = 31.8$ است، پس برای این داده‌ها دامنه‌ی تغییرات برابر است با:

$$R = \max - \min$$

$$= 31.8 - 6.2 = 25.6$$

بنابراین اگر طول رده‌ها را برابر بگیریم طول هر رده، C ، از فرمول زیر حاصل می‌شود.

$$c = \frac{R}{k} = \frac{۲۵/۶}{۷} = ۳/۶۵$$

در این مثال، برای سادگی، طول مشترک داده‌ها را برابر ۴ انتخاب خواهیم کرد. اگر بنابر توصیه‌های ذکر شده، حد پایین رده‌ی اول، یعنی کوچکترین مقداری را که بتوان در رده‌ی اول قرار گیرد برابر ۵ انتخاب کنیم، در این صورت می‌توانیم از ۷ رده‌ی ۸/۹-۵/۰، ۱۲/۹-۹/۰، ...، ۳۲/۹-۲۹/۰ استفاده کنیم. در این حالت که حدود به فرمت داده‌هاست، حد پایینی را با L^* و حد بالایی را با U^* نمادگذاری می‌کنیم. توجه می‌کنیم که طول هر رده ۴ است زیرا مثلاً رده‌ی اول ۸/۹-۵/۰ شامل کلیه‌ی مقادیر از ۴/۹۵ تا ۸/۹۵ است به این دلیل که داده‌های ما که در این مثال کمیت‌های پیوسته هستند، به نزدیکترین یک دهم تن گرد شده‌اند. به عنوان مثال داده‌ای که به صورت ۸/۳ ثبت شده است، در اصل مقداری بین ۸/۲۵ و ۸/۳۵ بوده و به نزدیکترین یک دهم تن گرد شده است. بنابر این حدود واقعی رده‌ی اول مقادیر ۴/۹۵ و ۸/۹۵، حدود واقعی رده‌ی دوم ۸/۹۵ و ۱۲/۹۵، ...، و حدود واقعی رده‌ی آخر مقادیر ۲۸/۹۵ و ۳۲/۹۵ است. حدود واقعی رده‌ها را حدود رده‌ها می‌نامیم که با حروف L و U به ترتیب برای حد پایینی و حد بالایی رده نمایش می‌دهیم.

جدول ۲-۶. جدول فراوانی اولیه‌ی میزان اکسید سولفور منتشره در هوا

$L^* - U^*$	شمار ش	f
۵/۰-۸/۹		۳
۹/۰-۱۲/۹		۱۰
۱۳/۰-۱۶/۹		۱۴
۱۷/۰-۲۰/۹		۲۵
۲۱/۰-۲۴/۹		۱۷
۲۵/۰-۲۸/۹		۹
۲۹/۰-۳۲/۹		۲
		$\sum_{i=1}^7 f_i = ۸۰$

اینک برای داده‌های مثال حاضر، جدولی تنظیم می‌کنیم و در اولین ستون آن حدود رده‌ها را درج می‌کنیم. ستون دیگری به نام ستون شمارش یا چوب‌خط (خط‌نشان) در نظر می‌گیریم. داده‌ها را یک‌به‌یک می‌خوانیم و در برابر رده‌ای که داده به آن تعلق دارد، یک خط عمودی به شکل | وارد می‌کنیم. برای سهولت در شمارش آنها، پس از هر

چهار علامت، خط موربی به صورت / به روی آنها می کشیم و هر چنین مجموعه ای به نشانه پنج مورد از داده هاست. حال به طوری که در جدول دیده می شود، تعداد «خط نشان ها» را شمرده در ستون سوم همین جدول وارد می کنیم و به این ترتیب فراوانی هر رده به دست می آید. برای $i = 1, 2, \dots, k$ فراوانی رده ی i م را با نماد f_i

$$\sum_{i=1}^k f_i = n \text{ نمایش می دهیم، بدیهی است که}$$

برای آنکه بتوان استفاده های گسترده تری از جدول فراوانی به عمل آورد، با افزودن ستون های دیگری، آن را کامل تر می کنیم. برای این منظور ستون دیگری به جدول فراوانی اضافه می کنیم که ستون **نماینده ی (مرکز) رده ها** نامیده می شود. برای آنکه نماینده ی هر رده را به دست آوریم، حد بالای هر رده را با حد پایین همان رده جمع و حاصل را بر دو تقسیم می کنیم. نماینده ی هر رده را می توان با محاسبه ی مجموع حدهای پایین و بالای هر رده و تقسیم حاصل بر ۲ نیز به دست آورد. برای داده های مثال ۲-۲ نماینده ی رده ها به ترتیب مقادیر $\frac{5,0+8,9}{2} = 6,95$ ، $\frac{9,0+12,9}{2} = 10,95$ ، ...، $\frac{29,0+32,9}{2} = 30,95$ هستند. نماینده ی رده ی i م را با نماد x_i نشان می دهیم. به طور کلی

$$x_i = \frac{L_i + U_i}{2}, \quad i = 1, 2, \dots, k$$

جدول ۲-۷. جدول فراوانی مربوط به میزان اکسید سولفور منتشره در هوا

$L^*_U^*$	L_U	x	f	$r \times 100$
۵,۰-۸,۹	۴,۹۵-۸,۹۵	۶,۹۵	۳	۳,۷۵
۹,۰-۱۲,۹	۸,۹۵-۱۲,۹۵	۱۰,۹۵	۱۰	۱۲,۵
۱۳,۰-۱۶,۹	۱۲,۹۵-۱۶,۹۵	۱۴,۹۵	۱۴	۱۷,۵
۱۷,۰-۲۰,۹	۱۶,۹۵-۲۰,۹۵	۱۸,۹۵	۲۵	۳۱,۲۵
۲۱,۰-۲۴,۹	۲۰,۹۵-۲۴,۹۵	۲۲,۹۵	۱۷	۲۱,۲۵
۲۵,۰-۲۸,۹	۲۴,۹۵-۲۸,۹۵	۲۶,۹۵	۹	۱۱,۲۵
۲۹,۰-۳۲,۹	۲۸,۹۵-۳۲,۹۵	۳۰,۹۵	۲	۲,۵
			$n = ۸۰$	۱۰۰

همچنین ستون دیگری به نام ستون فراوانی درصد به جدول فراوانی اضافه می‌کنیم. برای محاسبه‌ی فراوانی درصد هر رده، فراوانی رده را بر تعداد کل داده‌ها تقسیم و حاصل را در ۱۰۰ ضرب می‌کنیم. برای داده‌های مثال حاضر، فراوانی درصد رده‌ها به ترتیب اعداد $\frac{3 \times 100}{80} = 3,75$ ، $\frac{10 \times 100}{80} = 12,5$ ، ...، $\frac{2 \times 100}{80} = 2,5$ خواهند بود. به‌طور کلی

$$r_i = \frac{f_i}{n}, \quad i = 1, 2, \dots, k$$

داشتن فراوانی‌های درصد برای مقایسه‌ی دو مجموعه از داده‌ها بسیار مفید است. مثلاً اگر بخواهیم میزان اکسید سولفور منتشره از کارخانه‌ای را با میزان اکسید سولفور کارخانه‌ای دیگر مقایسه کنیم، حتماً باید از فراوانی‌های درصد استفاده کنیم. زیرا امکان دارد تعداد داده‌های مربوط به دو کارخانه متفاوت باشد. به علاوه برای این نوع مقایسه‌ها و موارد دیگر، می‌توان از **فراوانی تجمعی** نیز استفاده کرد که تعداد «داده‌های کمتر از» مقداری مشخص را نشان می‌دهد. برای آنکه تعداد داده‌های کمتر از حد پایین هر رده را حساب کنیم، فراوانی‌های تمام رده‌های پیش از آن رده را با هم جمع می‌کنیم. بدیهی است که فراوانی تجمعی آخرین رده برابر اندازه مشاهدات، n ، است. برای داده‌های مثال ۲-۳، جدول فراوانی تجمعی ۲-۸ را خواهیم داشت.

جدول ۲-۸. جدول فراوانی تجمعی مربوط به میزان اکسید سولفور منتشره در هوا

میزان اکسید سولفور بر حسب تن	F
کمتر از ۵,۰	$F_{\cdot} = f_{\cdot} = 0$
کمتر از ۹,۰	$F_1 = f_1 = 3$
کمتر از ۱۳,۰	$F_2 = f_1 + f_2 = 13$
کمتر از ۱۷,۰	$F_3 = f_1 + f_2 + f_3 = 27$
کمتر از ۲۱,۰	$F_4 = f_1 + f_2 + f_3 + f_4 = 52$
کمتر از ۲۵,۰	$F_5 = f_1 + f_2 + f_3 + f_4 + f_5 = 79$
کمتر از ۲۹,۰	$F_6 = f_1 + f_2 + f_3 + f_4 + f_5 + f_6 = 78$
کمتر از ۳۳,۰	$F_7 = f_1 + f_2 + f_3 + f_4 + f_5 + f_6 + f_7 = 80$

می‌توان جدول ۲-۸ را در جدول فراوانی ۲-۷، ادغام کرد و تمامی اطلاعات را یک‌جا در جدول ۲-۹ ارائه داد.

جدول ۲-۹. جدول کامل فراوانی مثال ۲-۳

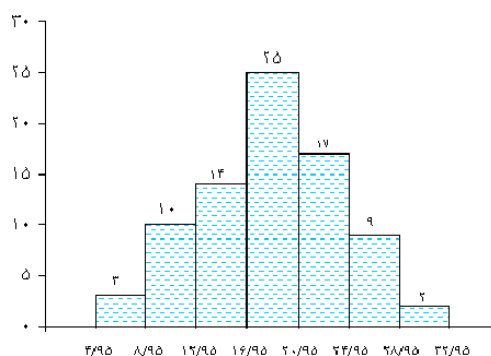
$L^*_U^*$	L_U	x	f	$r \times 100$	F
۵۰-۸۹	۴۹۵-۸۹۵	۶۹۵	۳	۳٫۷۵	۳
۹۰-۱۲۹	۸۹۵-۱۲۹۵	۱۰۹۵	۱۰	۱۲٫۵	۱۳
۱۳۰-۱۶۹	۱۲۹۵-۱۶۹۵	۱۴۹۵	۱۴	۱۷٫۵	۲۷
۱۷۰-۲۰۹	۱۶۹۵-۲۰۹۵	۱۸۹۵	۲۵	۳۱٫۲۵	۵۲
۲۱۰-۲۴۹	۲۰۹۵-۲۴۹۵	۲۲۹۵	۱۷	۲۱٫۲۵	۶۹
۲۵۰-۲۸۹	۲۴۹۵-۲۸۹۵	۲۶۹۵	۹	۱۱٫۲۵	۷۸
۲۹۰-۳۲۹	۲۸۹۵-۳۲۹۵	۳۰۹۵	۲	۲٫۵	۸۰
			$n = ۸۰$	۱۰۰	

ساخت جدول ۲-۹ را به وسیله نرم افزار «آر»، با تعیین ستون فراوانی، فراوانی درصد، و فراوانی تجمعی در فصل هفتم ارائه داده ایم.

۲-۳ توصیف داده ها به کمک نمودارها

نمودارها و تصاویر علاوه بر اینکه مشخصه های اصلی داده ها را به نحو بارزی آشکار می کنند، از لحاظ نظری نیز بسیار مفیدند. رایج ترین نمودار آماری، **بافت نگار** یا هیستوگرام نام دارد. برای رسم بافت نگار، از یک دستگاه محورهای مختصات متعامد استفاده می کنیم. حدود رده ها را روی محور افقی درج می کنیم و سپس مستطیل هایی بنا می کنیم به طوری که طول قاعده ی مستطیل مربوط به رده ی i م، طول آن رده و ارتفاع آن، فراوانی رده باشد.

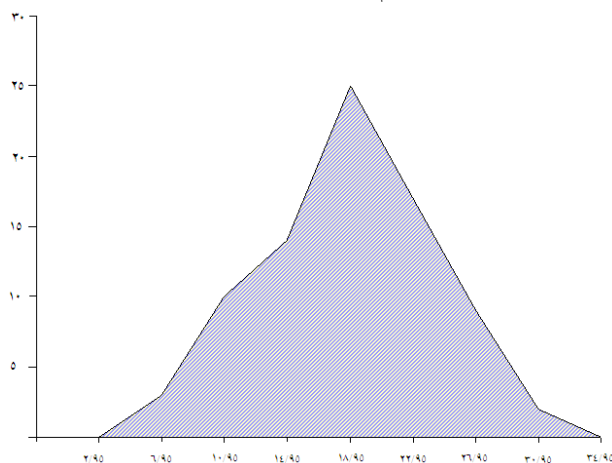
مثال ۲-۴. داده های مربوط به میزان اکسید سولفور منتشره در هوا را که در مثال ۲-۳ ارائه شد، یک بار دیگر در نظر می گیریم. برای رسم بافت نگار مربوط به داده ها، به طوری که گفتیم حدود رده ها را بر روی محور افقی ثبت می کنیم و مستطیل هایی بر روی این رده ها بنا می کنیم که ارتفاع آنها برابر با فراوانی رده ی مربوط باشد. توجه می کنیم که مقیاس روی دو محور را می توان متفاوت در نظر گرفت. بافت نگار مربوط به این مثال به صورت نمودار ۲-۱ است.



نمودار ۱-۲. نمودار بافت‌نگار میزان اکسید سولفور منتشره در هوا

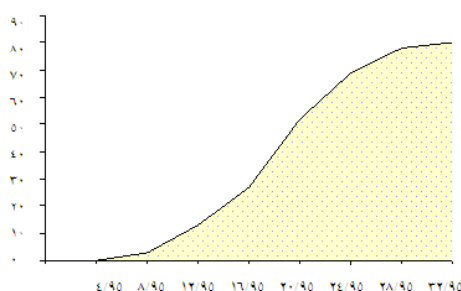
نمودار ۱-۲ و سایر نمودارهای این بخش را با نوشتن چند دستور نسبتاً ساده در فصل هفت می‌توان ترسیم کرد.

نمودار دیگری که کمتر مورد استفاده قرار می‌گیرد، **چندبر فراوانی** نام دارد. در این نمودار، مقادیر نماینده‌ی رده‌ها را روی محور افقی و فراوانی‌های مربوط را روی محور عمودی درج می‌کنیم و نقطه‌های حاصل را به کمک خط‌های راست به هم وصل می‌کنیم. برای آنکه نمودار حاصل را به محور افقی «ببندیم»، دو رده‌ی فرضی، یکی به مرکز ۲/۹۵ و دیگری به مرکز ۳۴/۹۵، و هر دو با فراوانی صفر، در نظر می‌گیریم، و از اولین نقطه‌ی به دست آمده به نقطه‌ی (۶/۹۵، ۳) و از نقطه‌ی (۳۰/۹۵، ۲) به آخرین نقطه، یعنی (۳۴/۹۵، ۰) وصل می‌کنیم تا نمودار ۲-۲ حاصل شود.



نمودار ۲-۲. چندبر فراوانی توزیع داده‌های میزان اکسید سولفور

اگر به همان صورت که در مورد چندبر فراوانی عمل کردیم، نموداری برای فراوانی تجمعی به دست آوریم، نمودار حاصل **اوجایو** نامیده می‌شود. برای رسم اوجایو مربوط به فراوانی تجمعی میزان اکسید سولفور منتشره در هوا، نقاطی را در دستگاه مختصات در نظر می‌گیریم که طول هر نقطه، حد بالای هر رده و عرض آن فراوانی تجمعی آن رده باشد. در این صورت نمودار ۲-۳ را به دست می‌آوریم.



نمودار ۲-۳. اوجایو توزیع داده‌های میزان اکسید سولفور

می‌توان به جای بافت‌نگار، از **نمودار میله‌ای** نیز استفاده کرد. برای این منظور عیناً مانند زمانی عمل می‌کنیم که می‌خواهیم بافت‌نگار را رسم کنیم، با این تفاوت که به جای استفاده از حدود رده‌ها، از نماینده‌ی رده‌ها استفاده می‌کنیم. با اینکه هم برای داده‌های پیوسته و هم گسسته ترسیم نمودارهای بافت‌نگار و میله‌ای انجام می‌شود، معمولاً در تنظیم یک گزارش تحقیقاتی رسم نمودار میله‌ای برای نمایش داده‌های گسسته و رسم نمودار بافت‌نگار برای داده‌های پیوسته رعایت می‌شود.

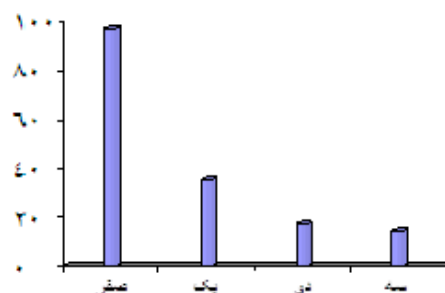
مثال ۲-۵. جدول ۲-۱۰ اطلاعات مربوط به تعداد روزهای مرخصی در یک ماه برای ۱۶۳ کارمند موسسه‌ای جمع‌آوری شده است.

جدول ۲-۱۰. اطلاعات مربوط به مثال ۲-۴

تعداد روزهای مرخصی	۰	۱	۲	۳
فراوانی	۹۷	۳۵	۱۷	۱۴

جدول و نمودارهای آماری ۳۱

نمودار ۲-۴ برای اطلاعات مثال ۲-۴ ترسیم شده است. زیرنویس هر رده با کلمات «صفر»، «یک»، «دو»، و «سه» مشخص شده تا این نمودار برای متغیرهای اسمی و ترتیبی نیز به کار می‌رود.



نمودار ۲-۴. نمودار میله‌ای داده‌های روزهای مرخصی در یک سازمان

نمودار دیگری که اغلب در روزنامه‌ها و گزارش‌های مختلف مشاهده می‌شود، **نمودار دایره‌ای (کلوچه‌ای)** نام دارد و عمدتاً برای ارائه‌ی فراوانی نسبی یا فراوانی درصد به کار می‌رود. برای رسم نمودار دایره‌ای، کل زاویه‌ی مرکزی 360° مربوط به یک دایره را به نسبت فراوانی‌های نسبی یا درصد تقسیم می‌کنیم و سپس قطاع‌هایی در دایره رسم می‌کنیم که زاویه‌ی مرکزی هریک از آنها برابر با اعداد به دست آمده باشد. اگر فراوانی نسبی رده‌ای r_i باشد، زاویه‌ی مرکزی قطاع مربوط به آن رده را برابر:

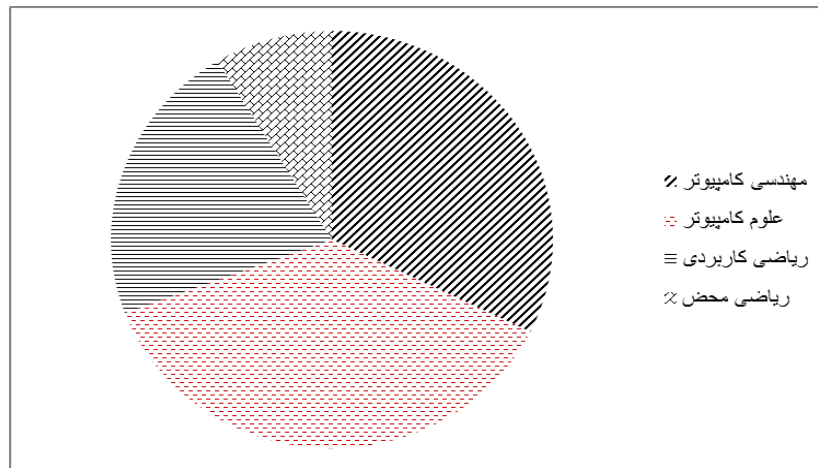
$$\alpha_i = 360^\circ \times r_i, \quad i = 1, 2, \dots, k$$

می‌گیریم.

جدول ۲-۱۱. یک مثال از توزیع رشته‌ای

عنوان رشته	f	r	α
مهندسی کامپیوتر	۴۱	۰/۳۲	۱۱۶
علوم کامپیوتر	۴۷	۰/۳۷	۱۳۳
ریاضی کاربردی	۲۸	۰/۲۲	۸۰
ریاضی محض	۱۱	۰/۰۹	۳۱

نمودار ۲-۵، نمودار دایره‌ای، برای اطلاعات جدول ۲-۱۱ که نمونه‌ای از دانشجویان دانشگاه پیام‌نور مرکز شیراز می‌باشد، ترسیم شده است.



نمودار ۲-۵. نمودار دایره‌ای برای داده‌های جدول ۲-۱۱

خلاصه‌ی فصل دوم

در این فصل، دو بخش مهم یکی ساخت جداول آماری و دیگری رسم برخی نمودارهای آماری آموزش داده شده است. مهم‌ترین مسئله در شروع این دو بخش، تعیین نوع متغیرها است. اگر متغیر تصادفی از نوع گسسته‌ای بود که تعداد مقادیری که اختیار می‌کند بزرگ نباشد، خود آن اعداد به عنوان رده معرفی می‌شود. برای متغیرهای پیوسته از دستورهای لازم جهت تشکیل رده‌های پیوسته، در جداول آماری استفاده می‌کنیم. نمودارهای بافت‌نگار، چندبر فراوانی، و اجایو معمولاً برای داده‌هایی که از یک صفت پیوسته اندازه‌گیری شده‌اند ترسیم می‌شود. نمودار میله‌ای و نمودارهای دایره‌ای برای متغیرهای گسسته و یا مشاهداتی که از روی مقیاس اسمی حاصل شده‌اند کاربرد دارد.

تمرین

۲-۱ داده‌های زیر تعداد مشتریان یک رستوران در یک ماه است.

۲۳	۲۸	۱۹	۱۴	۳۰	۴۴	۶۱
۱۴	۲۹	۱۸	۱۶	۲۰	۶۰	۶۶

جدول و نمودارهای آماری ۳۳

۶۷	۶۸	۲۳	۱۷	۲۰	۱۳	۲۶
۵۱	۶۱	۲۸	۲۸	۲۲	۲۰	۲۷
۲۷	۱۸					

- الف. نمودار ساقه و برگ را به صورت دستی تشکیل دهید.
- ب. بار دیگر به کمک نرم افزار، نمودار ساقه و برگ را رسم کنید.
- ۲-۲. داده های زیر اندازه ی قد ۱۰۰ پسر ۲۰ ساله در شهر شیراز است.

۱۷۲	۱۵۸	۱۶۲	۱۶۳	۱۷۲	۱۶۲	۱۶۳	۱۸۴	۱۶۶	۱۵۷
۱۸۴	۱۶۵	۱۵۴	۱۷۰	۱۶۹	۱۶۷	۱۷۰	۱۷۷	۱۶۴	۱۶۳
۱۷۰	۱۶۹	۱۵۹	۱۷۷	۱۶۲	۱۸۰	۱۶۵	۱۵۳	۱۷۴	۱۵۳
۱۵۰	۱۷۲	۱۸۲	۱۸۴	۱۷۲	۱۶۹	۱۷۲	۱۷۰	۱۶۹	۱۵۸
۱۶۴	۱۶۶	۱۷۴	۱۷۵	۱۷۶	۱۵۲	۱۸۴	۱۶۲	۱۷۹	۱۵۱
۱۷۱	۱۶۴	۱۶۲	۱۷۱	۱۷۵	۱۵۹	۱۷۰	۱۵۴	۱۷۷	۱۶۲
۱۷۸	۱۶۹	۱۵۱	۱۶۴	۱۷۴	۱۶۱	۱۵۰	۱۷۳	۱۷۹	۱۵۹
۱۷۲	۱۵۹	۱۶۵	۱۵۶	۱۶۹	۱۶۴	۱۶۴	۱۷۷	۱۷۳	۱۶۱
۱۷۷	۱۵۳	۱۷۲	۱۵۰	۱۶۶	۱۷۱	۱۷۱	۱۶۹	۱۶۵	۱۶۰
۱۵۱	۱۷۰	۱۵۶	۱۶۵	۱۵۴	۱۵۶	۱۷۰	۱۷۲	۱۶۵	۱۸۳

- الف. نمودار ساقه و برگ آن را رسم کنید.
- ب. با استفاده از روش تعیین تعداد رده ها (یکی از آنها)، جدول فراوانی آن را به دست آورید.
- ج. نمودار بافت نگار و چندبر فراوانی آن را رسم کنید.
- د. نمودارهای به دست آمده در قسمت ج را تفسیر کنید.
- ۲-۳. سازمان میراث فرهنگی، صنایع دستی و گردشگری استان زنجان در ایام نوروز سال ۱۳۹۴ گزارش عملکرد بازدید و نظارت مداوم از مکان های گردشگری و اقامتی را چنین اعلام نمود: ۳۳ مورد تذکر، ۱۲ مورد اخطار کتبی و ۲۸ مورد شکایت. نمودار مناسب برای توصیف داده های را در نظر گرفته آن را رسم کنید.
- ۲-۴. جدول ناقص زیر اطلاعات مربوط به تعداد روزهای لازم برای اتمام چهارچوب خانه های پیش ساخته، تنظیم شده است.

m	f	rf	F
۵	...	...	...
۶	۱۹	...	۲۷
۷	...	۰٫۳۵	...
۸	۱۵	۰٫۱۵	...
۹	...	...	۹۱
۱۰	...	...	...
		۱۰۰	۱٫۰۰

الف. جدول را تکمیل کنید.

ب. نمودار میله‌ای را رسم کنید.

ج. شرحی برای نمودار میله‌ای بنویسید.

۲-۵. داده‌های زیر مربوط به هزینه‌های حمل‌ونقل زمینی در یک سال شرکتی برحسب صدمیلیون ریال است. جدولی برمبنای پنج رده به‌طوری که شروع رده‌ی اول آن با پانزده باشد، ساخته و همچنین نمودار مناسب برای توصیف داده‌ها را رسم کنید.

۲۴	۲۱	۲۰	۳۱	۲۵
۴۱	۴۳	۳۲	۱۷	۲۵
۳۲	۳۰	۳۰	۲۶	۳۴
۳۶	۳۹	۲۷	۲۸	۳۱
۲۱	۱۹	۲۵	۲۶	۳۱

خودآزمایی چهارگزینه‌ای

۲-۱. برای مشاهدات ۱۲، ۱۴، ۱۷، ۱۹، ۲۲، ۲۳، ۳۲، ۳۵، ۳۸ نمودار ساقه و برگ از

چند عدد تک رقمی تشکیل می‌شود؟

الف. ۱۲ ب. ۹ ج. ۱۰ د. ۱۸

۲-۲. در فرمول تجربی $k = 1 + 3 \log(n)$ تعداد طبقات

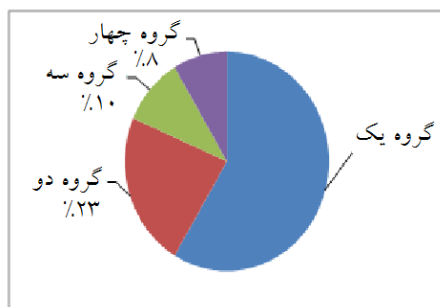
الف. با اندازه‌ی نمونه متناسب است.

ب. در نمونه‌های کمتر از ۱۰۰۰۰، بیش از ۳۵ است.

ج. کمتر از پنج رده انتخاب نمی شود.

د. می تواند بسته به هویت نوع داده ها تغییر داده شود.

۲- ۳. در نمودار دایره ای، مقدار زوایه ی مربوط به «گروه یک» کدام است؟



الف. ۵۸ ب. ۲۰۹ ج. ۲۷۰ د. ۷۰

۲- ۴. مشاهدات در جدول ناقص زیر خلاصه شده اند، مقدار f برابر کدام گزینه

است؟

رده ها	فراوانی	فراوانی نسبی	فراوانی تجمعی
۲۵_۴۵	a	d	g
۴۵_۶۵	۱۹	e	۲۷
۶۵_۸۵	b	۰٫۳۵	h
۸۵_۱۰۵	۱۵	۰٫۱۵	i
۱۰۵_۱۲۵	c	f	۹۱
۱۲۵_۱۴۵	k	j	l

الف. ۰٫۱۴ ب. ۰٫۰۸ ج. ۰٫۲۷ د. ۰٫۱۹

۲- ۵. مقدار h در جدول سؤال چهارگزینه ای ۲-۴ برابر کدام گزینه است؟

الف. ۷۷ ب. ۶۲ ج. ۵۴ د. ۴۶

۲- ۶. کدام نمودار برای داده ها با مقیاس فاصله ای به کار می رود؟

الف. بافت نگار ب. میله ای ج. دایره ای د. جعبه ای

۲- ۷. اگر ستون فراوانی را در دو ضرب کنیم، نسبت مساحت مستطیل ها نسبت به

مستطیل های اولیه ی همتای خود، در نمودار بافت نگار برابر کدام گزینه است؟

الف. ۲ ب. ۴ ج. ۱ د. ۰٫۵

۲-۸. مجموع مساحت مستطیل‌ها در نمودار بافت‌نگار است؟

- الف. برابر با یک
 ج. بیشتر از یک
 ب. کمتر از یک
 د. اگر طول رده‌ها برابر نباشد، کمتر از یک
- ۲-۹. اگر در جدول آماری، طول رده‌ها با هم برابر نباشند، آنگاه بر روی محور عمودی، چه مقداری به هر رده تخصیص می‌دهیم؟ (c_i را طول رده‌ی i م در نظر بگیرید.)

الف. $\frac{f_i}{c_i}$ ب. f_i ج. c_i د. $c_i f_i$

۲-۱۰. سطح زیر نمودار چندبر فراوانی است؟

- الف. یک
 ب. کمتر از یک
 ج. تقریباً یک
 د. بیش از یک
- ۲-۱۱. در نمودار ساقه و برگ

- الف. تصویری از وضع کلی داده‌ها را در اختیار می‌گذارد.
 ب. وقتی تعداد داده‌ها کم باشد اطلاعات بهتری ارائه می‌دهد.
 ج. نمی‌توان شاخه‌های بیشتری را رسم نمود.
 د. از چند ساقه می‌توان استفاده کرد.

۲-۱۲. توزیعی که برای داده‌های دسته‌بندی‌شده‌ی غیرعددی به کار می‌رود چه نام دارد؟

- الف. رسته‌ای
 ب. کمی
 ج. غیرعددی
 د. فراوانی
- ۲-۱۳. اگر بزرگترین داده برابر با ۸۹ و کوچکترین داده ۴۵ و تعداد طبقات ۵ باشد دامنه‌ی تغییرات چقدر است؟

الف. ۸/۸ ب. ۱۸ ج. ۴۴ د. ۹

۲-۱۴. اگر درصد افراد یک گروه در جامعه‌ای ۳۰٪ باشد، در نمودار دایره‌ای چه زاویه‌ای از دایره را به این گروه اختصاص می‌دهیم؟

الف. 30° ب. 120° ج. 108° د. 54°

فصل سوم

معیارهای مرکزی و پراکندگی

هدف کلی

یادگیری مفاهیم و محاسبه‌ی معیارهای مرکزی و پراکندگی برای اطلاعات به‌دست آمده از یک نمونه‌ی تصادفی

هدف‌های یادگیری

خواننده پس از یادگیری این فصل باید بتواند:

- انواع میانگین‌ها شامل حسابی، توافقی، هندسی، ریشه‌ای رتبه‌ی دو را محاسبه کند.
- ویژگی‌ها و تفاوت‌های بین انواع میانگین‌ها را بیان نماید.
- میانه را برای داده‌های خام و جدول آماری تعیین کند.
- برای هر داده‌ای، در صورت وجود، مد را تعیین کند.
- معیارهای پراکندگی، شامل دامنه، واریانس و انحراف معیار را محاسبه کند.
- میانگین و واریانس را از طریق کدگذاری به‌دست آورد.
- ویژگی‌های ضریب تغییرات را بیان کرده و آنرا به‌دست آورد.
- مفهوم نمره‌ی استاندارد را بیان کرده و آنرا محاسبه کند.
- قضیه‌ی چیبشف را بیان کرده و آنرا با محاسبه از جدول آماری مقایسه کند.
- چندک‌ها را بیان کند و انواع آنرا به‌دست آورد.
- دو معیار پراکندگی را که از روی چندک‌ها محاسبه می‌شود به‌دست آورد.

- نمودار جعبه‌ای را که براساس معیارهای گرایش مرکزی است را ترسیم کند.

مقدمه

در فصل دوم نحوه‌ی تلخیص داده‌ها، نحوه‌ی تشکیل جدول فراوانی، و شیوه‌ی توصیف داده‌ها را با رسم نمودارهای مختلف شرح دادیم. طبعاً این‌گونه نمودارها برای بیان چگونگی توزیع داده‌ها به قدر کافی گویا و اطلاع‌دهنده‌اند، اما جای ویژگی‌هایی از داده‌ها را که به کمک اعداد قابل توصیف است، نمی‌گیرند. منظور از توصیف عددی داده‌ها آن است که کمیت‌هایی عددی را از روی داده‌ها به دست آوریم که به بهترین نحو معرف داده‌ها باشند. این کمیت‌ها، که **معیار (شاخص)** نیز نامیده می‌شوند بر دو نوع‌اند: **معیارهای گرایش مرکزی** یا اندازه‌های مکانی، و **معیارهای پراکندگی**. در زیر به شرح هریک از انواع این اندازه‌ها می‌پردازیم.

۳-۱- معیارهای گرایش مرکزی

اغلب لازم است که مجموعه‌ای از داده‌ها را به کمک عدد واحدی بنمایانیم که به گونه‌ای، تا حد امکان، معرف تمام مجموعه‌ی داده‌ها باشد. اینکه عدد مزبور را چگونه انتخاب کنیم، بستگی به ویژگی خاصی دارد که می‌خواهیم به توصیف آن پردازیم. در مطالعه‌ای معین، ممکن است علاقه‌ی ما به مقادیر فرین (کمترین و بیشترین مقدار) در مجموعه داده‌ها باشد، در مطالعه‌ای دیگر ممکن است توجه ما به عددی باشد که تنها ۱۰ درصد داده‌ها از آن بیشترند، و نظایر آنها. چندین اندازه را در بخش‌های آتی مورد بحث قرار می‌دهیم که به گونه‌ای مرکز یا وسط مجموعه‌ای از داده‌ها را نشان می‌دهند و به همین دلیل آنها را معیارهای گرایش مرکزی می‌نامیم.

۳-۱-۱- میانگین

در بین انواع معیارهای گرایش مرکزی، مشهورتر از همه **میانگین حسابی** یا به اختصار **میانگین** است. میانگین مجموعه‌ای از مقادیر عبارت است از: مجموع آنها تقسیم بر تعدادشان. این همان چیزی است که کم‌وبیش از بدو ورود به دبستان با آن با نام معدل آشنایی داریم. برای آنکه معدل نمرات دانش‌آموزی را پیدا کنیم، مجموع نمرات وی را بر تعداد نمره‌هایش تقسیم می‌کنیم.

می‌توان با استفاده از نمادهای ریاضی، فرمولی کلی برای میانگین ارائه داد: فرض کنید n عدد که آنها را با حروف $x_1, x_2, \dots, x_n$ نشان می‌دهیم، داشته باشیم. اگر میانگین این اعداد را با $\bar{x}$ نشان‌دهیم آنگاه

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

می‌توان با استفاده از نماد Σ ، این فرمول را حتی به صورتی ساده‌تر نوشت:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i. \quad (1-3)$$

اگر x_i ها مقادیر یک جامعه‌ی آماری N عضوی باشند، آنگاه میانگین آنها را با حرف μ نمایش می‌دهیم.

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i.$$

مثال ۳-۱. حجم آب حاصل از ریزش در دریای خزر بر حسب کیلومتر مکعب در پنج سال به قرار زیر بوده است:

$$76,91 \quad 65,23 \quad 91,87 \quad 81,88 \quad 79,76$$

میانگین حجم آب حاصل از ریزش در این پنج سال چقدر است؟

حل. بنابر فرمول میانگین داریم:

$$\bar{x} = \frac{76,91 + 65,23 + 91,87 + 81,88 + 79,76}{5} = 79,13.$$

تمامی محاسبات معیارهای گرایش مرکزی و معیارهای پراکندگی به وسیله‌ی نرم‌افزار «آر» در بخش ۷-۳ به تفصیل شرح داده شده است.

باید توجه کنیم که فرمول میانگین که در فرمول (۱-۳) ارائه شد، برای داده‌هایی است که هنوز در جدول فراوانی گروه‌بندی نشده‌اند. اگر داده‌ها را در جدول فراوانی تلخیص کرده باشیم و داده‌های اصلی را در اختیار نداشته باشیم، در این صورت نحوه‌ی محاسبه‌ی میانگین متفاوت است. فرض کنید که نماینده یا مرکز رده‌ها را با حروف $x_1, x_2, \dots, x_k$ و فراوانی‌های متناظر را به ترتیب $f_1, f_2, \dots, f_k$ نشان دهیم. در این صورت فرمول زیر برای میانگین داده‌های گروه‌بندی شده به دست می‌آید:

$$\bar{x} = \frac{f_1 x_1 + f_2 x_2 + \dots + f_k x_k}{n},$$

یا به طور خلاصه:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k f_i x_i \quad (2-3)$$

بدیهی است به علت فشردگی اطلاعات در جداول آماری و استفاده از نماینده‌های رده‌ها، مقدار میانگین با داده‌های گروه‌بندی نشده قدری با مقدار عددی حاصل از فرمول (۲-۳) متفاوت خواهد شد. در حقیقت فرمول (۲-۳) برآورد یا تخمینی از مقدار واقعی میانگین جامعه خواهد بود.

مثال ۲-۳. میانگین داده‌های مربوط به میزان اکسید سولفور منتشره در هوا در ۸۰ روز را که در مثال (۲-۳) ارائه شده است، پیدا کنید.

جدول ۱-۳. جدول فراوانی برای محاسبه‌ی میانگین

L_U	x	f	f × x
۴٫۹۵ _ ۸٫۹۵	۶٫۹۵	۳	۲۰٫۸۵
۸٫۹۵ _ ۱۲٫۹۵	۱۰٫۹۵	۱۰	۱۰۹٫۵۰
۱۲٫۹۵ _ ۱۶٫۹۵	۱۴٫۹۵	۱۴	۲۰۹٫۳۰
۱۶٫۹۵ _ ۲۰٫۹۵	۱۸٫۹۵	۲۵	۴۷۳٫۷۵
۲۰٫۹۵ _ ۲۴٫۹۵	۲۲٫۹۵	۱۷	۳۹۰٫۱۵
۲۴٫۹۵ _ ۲۸٫۹۵	۲۶٫۹۵	۹	۲۴۲٫۵۵
۲۸٫۹۵ _ ۳۲٫۹۵	۳۰٫۹۵	۲	۶۱٫۹۰
		$\sum_{i=1}^v f_i = ۸۰$	$\sum_{i=1}^v f_i x_i = ۱۵۰۸$

حل. گرچه برای محاسبه‌ی میانگین داده‌های رده‌بندی شده تنها به نماینده‌های رده‌ها و فراوانی هر رده نیاز داریم، برای استفاده‌های آتی، جدول را به طور کامل مجدداً در جدول ۱-۳ می‌آوریم. به طوری که ملاحظه می‌شود علاوه بر اطلاعات موجود در جدول فراوانی پیشین، ستونی را اضافه کرده‌ایم که در آن نماینده‌ی هر رده در فراوانی نظیر آن ضرب و حاصل در این ستون درج شده است. مجموع اعداد واقع در این ستون همان $\sum_{i=1}^v f_i x_i = ۱۵۰۸$ است. چون در مورد این جدول فراوانی $\sum_{i=1}^v f_i x_i = ۱۵۰۸$ ، پس

$$\bar{x} = \frac{1508}{80} = 18,85.$$

مجدداً تأکید می‌کنیم که این میانگین با میانگین واقعی داده‌ها کمی تفاوت دارد و دلیل آن این است که مثلاً در رده‌ی اول به‌جای هر سه قلم از داده‌ها که مقادیری بین ۴/۹۵ تا ۸/۹۵ دارند، عدد ۶/۹۵، یعنی نماینده‌ی این رده را قرار دادیم، در حالی که میانگین داده‌های خام برابر با ۱۸/۹۳ است.

میانگین، متداول‌ترین اندازه‌ی گرایش مرکزی است. می‌توان دلایل زیر را برای مورد توجه بودن آن برشمرد:

۱. می‌توان آن را برای هر مجموعه‌ای از داده‌های عددی محاسبه کرد.
۲. مجموعه‌ای از داده‌های عددی یک و فقط یک میانگین دارد، و به‌عبارت دیگر منحصر به فرد است.
۳. می‌توان اعمال عددی بیشتری روی آن انجام داد (مثلاً اگر میانگین چند مجموعه از داده‌ها را داشته باشیم، به‌سادگی می‌توانیم میانگین مجموعه‌ای از داده‌ها را که از تلفیق کردن کلیه‌ی داده‌ها یک‌جا به دست آمده است حساب کنیم).
۴. میانگین نسبتاً قابل اطمینان است به این معنی که اگر چند نمونه‌ی مختلف را از جامعه‌ای استخراج کنیم، میانگین‌ها تفاوت خیلی زیادی با هم ندارند.
۵. در محاسبه‌ی میانگین، همه‌ی داده‌ها به حساب می‌آیند.
۶. با اعمال تبدیل زیر بر روی داده‌ها

$$y_i = ax_i + b; \quad i = 1, 2, \dots, n$$

رابطه‌ی زیر برقرار است:

$$\bar{y} = a\bar{x} + b$$

که در آن a و b اعداد حقیقی ثابتی هستند. خصوصاً با انتخاب $b = -\bar{x}$ و $a = 1$ صفرشدن میانگین y_i ‌ها، $\bar{y} = 0$ ، نتیجه می‌شود.

۳-۱-۲ میانه

در بخش ۳-۱-۱ یکی از امتیازات میانگین را در به حساب آوردن کلیه‌ی داده‌ها دانستیم. منظور این است که برای محاسبه‌ی میانگین باید همه‌ی داده‌ها را با هم جمع و

بر تعداد کل آنها تقسیم کرد. این امر گرچه ممکن است از جهتی امتیاز به حساب آید، در مواردی نیز ممکن است نقطه‌ی ضعف محسوب شود. برای مثال فرض کنید اعداد زیر حقوق ماهانه‌ی (برحسب هزار تومان) اعضای شرکتی را نشان دهد که پنج عضو دارد:

$$۱۲۱۵ \quad ۱۱۹۵ \quad ۱۲۴۵ \quad ۱۱۸۵ \quad ۱۹۸۵$$

اگر میانگین حقوق را با $\bar{x}$ نشان دهیم، آنگاه:

$$\bar{x} = \frac{۱۲۱۵ + ۱۱۹۵ + ۱۲۴۵ + ۱۱۸۵ + ۱۹۸۵}{۵} = ۱۳۶۵.$$

جایگاه میانگین نسبت به پنج مشاهده در زیر نمایش داده می‌شود.

$$۱۱۸۵ \quad ۱۱۹۵ \quad ۱۲۱۵ \quad ۱۲۴۵ \quad \downarrow \quad ۱۹۸۵$$

$$\bar{x} = ۱۳۶۵$$

به‌طوری که ملاحظه می‌کنیم، رقم ۱۹۸۵ هزار تومان حقوق یک فرد در مقایسه با حقوق بقیه‌ی افراد موجب شده‌است که میانگین حقوق، ۱۳۶۵ هزار تومان از کار درآید که نماینده‌ی واقعی حقوق‌های کارمندان این شرکت نیست. دلیل این کار، همان‌طور که گفتیم، به حساب آمدن همه‌ی اعضای مجموعه‌ی داده‌ها و در نتیجه تأثیر غیرعادی اقلامی از داده‌هاست که شاید با بقیه‌ی داده‌ها هماهنگ به نظر نرسند (درحقیقت داده‌ی ۱۹۸۵ یک داده‌ی پرت یا افراطی محسوب می‌شود که میانگین داده‌ها را به سمت خود می‌کشد). در چنین مواردی می‌توان اندازه گرایش مرکزی دیگری، به نام **میانه** را محاسبه کرد که به نوبه خود «وسط» یا «مرکز» داده‌ها را نشان می‌دهد. برای تعیین میانه در داده‌های گسسته (همچنین در داده‌های جدول‌بندی‌نشده)، داده‌ها را به ترتیب صعودی یا غیرنزولی مرتب می‌کنیم. اگر تعداد داده‌ها فرد باشد، داده‌ی وسطی میانه است. اگر تعداد داده‌ها زوج باشد، میانگین دو داده‌ی وسطی را میانه به حساب می‌آوریم. به عبارت دیگر اگر تعداد داده‌ها را وقتی فرد است با $n = 2k + 1$ نشان دهیم، آنگاه پس از مرتب کردن داده‌ها به ترتیب صعودی یا نزولی، k تا از داده‌ها در سمت چپ داده‌ی $(k + 1)$ م میانه است. اگر تعداد داده‌ها زوج باشد و $n = 2k$ ، در این صورت پس از مرتب کردن داده‌ها به صورت صعودی یا نزولی $k - 1$ تا از داده‌ها در سمت چپ داده‌ی k م و $k - 1$ تا از داده‌ها در سمت راست آن قرار می‌گیرند و لذا داده‌ی $(k + 1)$ م میانه به صورت صعودی یا نزولی $k - 1$ تا از داده‌ها در سمت چپ داده‌ی k م و $k - 1$ تا از داده‌ها در سمت راست داده‌ی $(k + 1)$ م قرار دارند و لذا میانه برابر با

میانگین حسابی داده‌ی k م و $(k+1)$ م است. معمولاً میانه‌ی داده‌ها را با Md نشان می‌دهیم. به‌طور کلی برای n عدد که مرتب‌شده‌ی آنها از کوچک به بزرگ به‌صورت $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ نمایش داده می‌شود، میانه عبارت است از:

$$Md = \begin{cases} x_{(\frac{n+1}{2})} & \text{فرد } n \\ \frac{1}{2}(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}) & \text{زوج } n \end{cases} \quad (3-3)$$

به‌عنوان مثال اگر حقوق پنج عضو شرکت را که در بالا ذکر شد، به ترتیب صعودی مرتب کنیم، ارقام زیر را خواهیم داشت:

۱۱۸۵ ۱۱۹۵ ۱۲۱۵ ۱۲۴۵ ۱۹۸۵

چون تعداد داده‌ها فرد است، لذا میانه‌ی داده‌ها برابر ۱۲۱۵ است که مناسب‌تر از $\bar{X} = ۱۳۶۵$ برای معرفی مرکزیت داده‌هاست.

مثال ۲-۳. مدت زمان (بر حسب دقیقه) تأخیر ورود به اداره‌ای برای ده کارمند به شرح ذیل ثبت شده است:

۰ ۱۱ ۴/۵ ۱۳ ۷/۸ ۳/۲ ۴۶ ۱۰ ۰ ۲/۵

میانه‌ی این مجموعه‌ی داده‌ها را به‌دست آورید.

حل. تعداد این داده‌ها $n=۱۰$ و بنابراین زوج است. اگر این اعداد را به‌ترتیب صعودی مرتب کنیم، مجموعه‌ی اعداد زیر را خواهیم داشت.

۰ ۰ ۲/۵ ۳/۲ ۴/۵ ۷/۸ ۱۰ ۱۱ ۱۳ ۴۶

در اینجا دو عدد $x_{(5)} = ۴/۵$ و $x_{(6)} = ۷/۸$ در وسط داده‌ها قرارگرفته‌اند. در نتیجه میانه‌ی داده‌ها به‌صورت زیر به دست می‌آید:

$$Md = \frac{1}{2}(x_{(5)} + x_{(6)}) = \frac{1}{2}(۴/۵ + ۷/۸) = ۶/۱۵.$$

اگر تعداد داده‌ها زیاد باشد، مرتب‌کردن داده‌ها به‌ترتیب صعودی یا نزولی کار چندان آسانی نیست، البته می‌توانیم نمایش ساقه و برگ داده‌ها را تهیه و مرکز داده‌ها را به‌دست آوریم. اما توجه داریم که معمولاً برای مقاصد دیگر هم، داده‌ها را معمولاً در

جدول توزیع فراوانی گروه‌بندی می‌کنیم، بنابر این لازم است روش تعیین میانه‌ی داده‌های جدول‌بندی شده را نیز بدانیم.

وقتی مجموعه‌ای از داده‌ها را در جدول فراوانی خلاصه کرده باشیم، پیدا کردن مقدار دقیق میانه امکان‌پذیر نیست، زیرا مقادیر واقعی داده‌هایی را که در رده‌ها قرار گرفته‌اند، در اختیار نداریم، اما می‌توان رده‌ای را که میانه در آن قرار گرفته است، پیدا کرد. بنابراین در این حالت میانه را به صورت دیگری تعریف می‌کنیم: میانه‌ی مجموعه‌ای از داده‌های رده‌بندی شده، عددی است که نصف مساحت مستطیل‌های بافت‌نگار در سمت چپ آن و نصف دیگر در سمت راست آن قرار دارد. این تعریف، معادل با این فرض است که مقادیر داده‌هایی که در رده‌ی حاوی میانه قرار دارند، به‌طور یکنواخت، یعنی با فاصله‌های برابر از هم، توزیع شده‌اند.

برای پیدا کردن خط تقسیم بین دو نیمه‌ی بافت‌نگار (که هر نیمه نماینده‌ی نصف داده‌های رده‌بندی، یعنی $n/2$ ، است)، باید با شروع از سمت چپ (یا راست) به طرف مرکز بافت‌نگار پیش رویم و به نقطه‌ای برسیم که $n/2$ داده‌ها در چپ (یا راست) آن قرار دارند. جزئیات بیشتر را در مثال زیر شرح می‌دهیم.

مثال ۳-۳. داده‌های مربوط به میزان اکسیدسولفور مثال ۲-۳ را که بافت‌نگار آن در نمودار ۲-۱ رسم شده است، در نظر می‌گیریم. برای این داده‌ها $n = 80$ و بنابراین $n/2 = 40$. پس باید مثلاً از سمت چپ (نقطه‌ی $4/95$) شروع کنیم و فراوانی‌های رده‌ها را با هم جمع کنیم تا به رده‌ای برسیم که میانه در آن قرار دارد. در این مثال،

$$3 + 10 + 14 = 27 < 40 < 3 + 10 + 14 + 25 = 52$$

پس میانه در رده‌ای با کرانه‌های $16/95$ و $20/95$ قرار دارد. تعداد $3 + 10 + 14 = 27$ تا از داده‌ها در سمت چپ نقطه‌ی $16/95$ قرار دارند، طول این رده (مانند همه‌ی رده‌ها در این مثال) ۴ است و ۲۵ مورد از داده‌ها در رده‌ی با کرانه‌های $16/95$ و $20/95$ قرار دارند. چون طول رده ۴ است، پس به هر داده $\frac{4}{25}$ از این طول اختصاص می‌یابد. اما $40 - 27 = 13$ ، پس باید $13 \times \frac{4}{25}$ را به عدد $16/95$ اضافه کنیم تا نقطه‌ای به دست آید

که ۴۰ تا از داده‌ها از آن کمترند و ۴۰ مورد دیگر از داده‌ها از آن بیشترند. بنابراین میانه‌ی داده‌های مربوط به این مثال، برابر است با:

$$Md = 16,95 + 13 \times \frac{4}{25} = 19,03.$$

در حالت کلی، محاسبه‌ی مقدار میانه برای داده‌های پیوسته (که از مقیاس فاصله‌ای و نسبتی به دست آمده‌اند) از فرمول (۳-۴) انجام می‌شود.

$$Md = L_k + \frac{\frac{n}{2} - F_{k-1}}{f_k} \times c. \quad (3-4)$$

که در آن L_k کران پایین طبقه‌ی میانه، F_{k-1} فراوانی تجمعی طبقه‌ی ماقبل طبقه‌ی میانه، f_k فراوانی طبقه‌ی میانه، c طول رده‌ی میانه است.

مثال ۳-۴. جدول فراوانی زیر، توزیع سنی کارگرانی را که در یک کارخانه‌ی صنعتی به کار اشتغال دارند، نشان می‌دهد. میانه‌ی این توزیع را پیدا کنید.

L_U	f	F
۱۹,۵ - ۲۴,۵	۱۲۹	۱۲۹
۲۴,۵ - ۲۹,۵	۲۲۱	۳۵۰
۲۹,۵ - ۳۴,۵	۳۱۰	۶۶۰
۳۴,۵ - ۳۹,۵	۱۶۳	۸۲۳
۳۹,۵ - ۴۴,۵	۱۰۵	۹۲۸
۴۴,۵ - ۴۹,۵	۶۲	۹۹۰
۴۹,۵ - ۵۴,۵	۱۰	$n = 1000$
	$\sum_i f = 1000$	

رده‌ی حاوی $Md \leftarrow$

حل. به طوری که ملاحظه می‌کنیم، در این جدول $n = 1000$ ، پس $\frac{n}{2} = 500$. با توجه به اینکه

$$F_{k-1} = 350 < 500 \leq 350 + 310 = F_k$$

پس میانه در رده‌ی سوم قرار دارد. بنابراین چون

$$\frac{n}{2} - F_{k-1} = 500 - 350 = 150, \quad c = 5, \quad f_r = 310,$$

$$Md = 29,5 + \frac{150}{310} \times 5 = 31,92.$$

۳-۱-۳ مُد

یکی دیگر از معیارهای گرایش مرکزی، که «مرکز» مجموعه‌ای از داده‌ها را توصیف می‌کند، مُد است. مُد مجموعه‌ای از داده‌ها، مقداری از داده‌هاست که بیشترین فراوانی را داشته باشد. می‌توان مزیت مد را در این دانست که:

الف. نیازی به محاسبه ندارد.

ب. می‌توان آن‌را برای داده‌های کیفی نیز تعیین کرد.

مثال ۳-۵. از ۱۰ دانشجو سؤال شده است که به کدام ورزش علاقه دارند و پاسخ‌های زیر دریافت شده است:

فوتبال، فوتبال، والیبال، فوتبال، بسکتبال، شطرنج، فوتبال، والیبال، والیبال، فوتبال.

ملاحظه می‌کنیم که فراوانی فوتبال پنج است و لذا مُد این مجموعه داده‌ها، فوتبال است.

مثال ۳-۶. نمره‌ی نهایی ۲۰ دانشجو در درس ریاضی عمومی به شرح زیر است. مد این داده‌ها را پیدا کنید.

۱۱,۰ ۱۵,۰ ۱۲,۰ ۱۲,۰ ۱۴,۰ ۷,۵ ۱۷,۰ ۸,۵ ۱۲,۰ ۱۴,۵

۷,۵ ۱۲,۰ ۸,۰ ۹,۰ ۱۴,۰ ۵,۵ ۷,۰ ۱۲,۰ ۱۶,۰ ۱۲,۰

با اندک واریسی ملاحظه می‌شود که تعداد دفعات تکرار یا فراوانی ۱۲، عدد ۶ است، لذا مُد این مجموعه‌ی داده‌ها ۱۲ است.

اگر داده‌ها در جدول فراوانی درج شده باشند، مجدداً باید به تعیین مُد تقریبی اکتفا کنیم. معمولاً فرمول‌هایی برای محاسبه‌ی مُد داده‌های رده‌بندی‌شده ارائه می‌شود. برای سهولت مُد دسته‌ای از داده‌های رده‌بندی‌شده را نماینده یا مرکز رده‌ای تعریف می‌کنیم

که بیشترین فراوانی را داشته باشد. در نمودارهای بافت‌نگار، مستطیلی که بزرگترین طول را داشته باشد، نماینده‌ی رده‌ی مربوط، نمایانگر مُد خواهد بود. در اکثر مواقع، برای داده‌های پیوسته، تأثیر فراوانی رده‌های قبل و بعد از رده‌ی مد را نمی‌توان نادیده گرفت. فرمول (۳-۵) که یک اثبات هندسی دارد، برای محاسبه‌ی مُد در داده‌های پیوسته به‌کار می‌رود. (بهبودیان، ۱۳۹۰)

$$Mo = L_m + \frac{d_1}{d_1 + d_2} \times c \quad (3-5)$$

که در آن L_m حد پایینی طبقه‌ای است که دارای بیشترین فراوانی است، d_1 اختلاف فراوانی رده‌ی مد با فراوانی رده‌ی قبل، d_2 اختلاف فراوانی رده‌ی مد با فراوانی رده‌ی بعد، c طول طبقه‌ی مد است.

مثال ۳-۷. در مثال ۲-۳، داده‌های مربوط به میزان اکسید سولفور، بیشترین فراوانی ۲۵، مربوط به رده‌ای است که کرانه‌های آن اعداد ۱۶٫۹۵ و ۲۰٫۹۵ است، پس مد این مجموعه‌ی داده‌ها به‌صورت تقریبی، عدد $Mo = ۱۸٫۹۵$ ، یعنی مرکز همین رده است. از طریق فرمول (۳-۵)، بدین صورت محاسبه می‌شود.

$$Mo = ۱۶٫۹۵ + \frac{(۲۵-۱۴)}{(۲۵-۱۴)+(۲۵-۱۷)} \times ۴ = ۱۹٫۲۷.$$

۳-۱-۴ دیگر معیارهای گرایش مرکزی

میانگین، میانه، و مد هر یک به‌گونه‌ای «مرکز» توزیع را نمایان می‌کنند. با این‌حال برای برخی داده‌ها به انواع دیگر اندازه‌های گرایش مرکزی نیاز داریم. در زیر دو مورد از این اندازه‌ها را مورد بحث قرار داده و مثال‌هایی از کاربرد آنها ارائه می‌کنیم.

میانگین توافقی (همساز) مجموعه‌ای از داده‌ها معکوس میانگین حسابی معکوس‌های داده‌هاست. به عبارت دیگر اگر داده‌ها را با $x_1, x_2, \dots, x_n$ و میانگین توافقی را با $\bar{x}_H$ نشان دهیم، آنگاه:

$$\frac{1}{\bar{x}_H} = \frac{1}{n} \sum_{i=1}^n \frac{1}{x_i}, \quad (3-6)$$

یا

$$\bar{x}_H = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}.$$

واضح است شرط $x_i \neq 0$ به ازای $i=1,2,\dots,n$ الزامی است. برای داده‌های رده‌بندی شده، فرمول محاسبه‌ی میانگین توافقی به صورت (۳-۷) در می‌آید.

$$\bar{x}_H = \frac{n}{\sum_{i=1}^n \frac{1}{x_i} f_i}, \quad (۳-۷)$$

که در آن x_i مرکز رده‌ی i م و f_i فراوانی این رده است. معمولاً وقتی مجموعه‌ی داده‌ها دارای واحدهای اندازه‌گیری ترکیبی است (مثلاً کیلومتر در ساعت)، از میانگین توافقی استفاده می‌شود.

مثال ۳-۸. سرعت خودرویی در ساعت اول ۴۸ کیلومتر در ساعت و در ساعت دوم ۸۰ کیلومتر در ساعت است، در این صورت میانگین توافقی

$$\bar{x}_H = 60 = 2 \left(\frac{1}{48} + \frac{1}{80} \right)^{-1}$$

محاسبه می‌شود.

مثال ۳-۹. سه کارگر به ترتیب دیواری به بلندی یک متر را در $\frac{1}{4}$ ، $\frac{1}{3}$ و $\frac{1}{2}$ ساعت می‌سازند. اگر این سه کارگر با هم کار کنند چقدر طول می‌کشد تا یک متر دیوار بنا شود؟

حل. پاسخ این سؤال عبارت است $\bar{x}_H = \frac{3}{\frac{1}{2} + \frac{1}{3} + \frac{1}{4}} = \frac{1}{\frac{1}{2} + \frac{1}{3} + \frac{1}{4}}$. لازم به ذکر است که در این مثال میانگین $\bar{x} = \frac{13}{12}$ نشان‌دهنده‌ی متوسط سرعت بالا بردن یک متر دیوار برای یک نفر است.

میانگین هندسی n عدد مثبت، ریشه‌ی n م حاصل ضرب آنهاست. به عبارت دیگر اگر عددهای مثبت را با $x_1, x_2, \dots, x_n$ و میانگین هندسی را با $\bar{x}_G$ نشان دهیم، آنگاه

$$\bar{x}_G = (x_1 x_2 \cdots x_n)^{\frac{1}{n}} = \sqrt[n]{x_1 x_2 \cdots x_n}. \quad (۳-۸)$$

با گرفتن لگاریتم از طرفین رابطه‌ی اخیر داریم:

$$\begin{aligned}\log(\bar{x}_G) &= \frac{1}{n} (\log(x_1) + \log(x_2) + \dots + \log(x_n)) \\ &= \frac{1}{n} \sum_{i=1}^n \log(x_i).\end{aligned}$$

پس میانگین هندسی، آنتی لگاریتم میانگین حسابی لگاریتم‌های داده‌هاست. اگر داده‌ها در یک جدول فراوانی تلخیص شده باشند، برای محاسبه‌ی میانگین هندسی، تعریف (۳-۹) را به کار می‌بریم.

$$\log(\bar{x}_G) = \frac{1}{n} \sum_{i=1}^k f_i \log(x_i) \quad (۳-۹)$$

که در آن x_i مرکز رده‌ی i م و f_i فراوانی این رده است. میانگین هندسی در پیدا کردن متوسط آهنگ‌های تغییر، مثلاً در جمعیت و غیره یا محاسبه‌ی میانگین نسبت تغییرات مانند افزایش قیمت‌ها و غیره، مفید است.

مثال ۳-۱۰. فرض کنید که جمعیت یک روستا ۱۰۰۰ نفر و آهنگ سالانه‌ی رشد جمعیت در طول چهار سال ۸٪ باشد. جمعیت این روستا در پایان این چهار سال و میانگین هندسی جمعیت پنج سال چقدر خواهد بود؟
حل. جمعیت این روستا در شروع و پایان هریک از این چهار سال، به ترتیب عبارت خواهد بود از:

$$\begin{aligned}& ۱۰۰۰, \quad ۱۰۰۰ \times (۱/۰۸)^1, \quad ۱۰۰۰ \times (۱/۰۸)^2, \quad ۱۰۰۰ \times (۱/۰۸)^3, \quad ۱۰۰۰ \times (۱/۰۸)^4 \\ & \text{میانگین هندسی این پنج عدد، ریشه‌ی پنجم حاصل ضرب این اعداد است، پس:} \\ & \bar{x}_G = \left(۱۰۰۰ \times ۱۰۰۰ \times (۱/۰۸)^1 \times ۱۰۰۰ \times (۱/۰۸)^2 \times ۱۰۰۰ \times (۱/۰۸)^3 \times ۱۰۰۰ \times (۱/۰۸)^4 \right)^{\frac{1}{5}} \\ & = \left((۱۰۰۰)^5 \times (۱/۰۸)^{1+2+3+4} \right)^{\frac{1}{5}} \\ & = ۱۰۰۰ \times (۱/۰۸)^2 = ۱۱۶۶.\end{aligned}$$

محاسبه‌ی میانگین هندسی و میانگین توافقی با نرم‌افزار نیازی به عملیات فوق ندارد و صرفاً با استفاده از فرمول مربوطه قابل انجام است. شرح دستورات لازم برای انجام محاسبات مربوطه با نرم‌افزار، در بخش ۷-۳ آمده است.

میانگین ریشه‌ای رتبه دو n عدد برابر با ریشه‌ی دوم، میانگین توان دوم آنهاست. به عبارت دیگر اگر عددها را با $x_1, x_2, \dots, x_n$ و میانگین ریشه‌ای رتبه دو را با M_2 نشان دهیم، آنگاه:

$$M_2 = \left(\frac{1}{n} \sum_{i=1}^n x_i^2 \right)^{\frac{1}{2}}. \quad (3-10)$$

اگر داده‌ها در یک جدول فراوانی تلخیص شده باشند، برای محاسبه‌ی میانگین ریشه‌ای رتبه دو، تعریف (3-11) را به کار می‌بریم.

$$M_2 = \left(\frac{1}{n} \sum_{i=1}^k f_i x_i^2 \right)^{\frac{1}{2}}. \quad (3-11)$$

که در آن m_i مرکز رده‌ی i ام و f_i فراوانی این رده است. در مشاهدات مثبت، رابطه‌ای بین میانگین‌های حسابی، هندسی، توافقی و ریشه‌ای رتبه دو به شکل (3-12) برقرار است.

$$M_2 \geq \bar{x} \geq \bar{x}_G \geq \bar{x}_H. \quad (3-12)$$

که به‌ازای $x_1 = x_2 = \dots = x_n$ برابری برقرار می‌شود. راه‌حل‌های متنوعی برای اثبات نابرابری (3-12) ارائه شده که در یکی از آنها با فرض مثبت بودن x_i ها، استفاده از

صعودی بودن تابع $M_j = \left(\frac{1}{n} \sum_{i=1}^k f_i x_i^j \right)^{\frac{1}{j}}$ برحسب j استفاده می‌شود. با قراردادن j برابر ۱- و ۰ و ۱ و ۲ در میانگین ریشه رتبه‌ی j ، نابرابری‌های (3-12) به‌دست می‌آید. توجه دارید که به‌ازای $j \neq 0$ داریم:

$$\lim_{j \rightarrow \infty} M_j = \bar{x}_G.$$

۳-۲ معیارهای پراکندگی

یکی از مهم‌ترین مشخصه‌های بیشتر داده‌ها، آن است که مقادیر داده‌ها همانند نیستند. در واقع، میزان ناهمانند بودن، یا تغییرات بین داده‌ها اهمیتی اساسی در علم آمار دارد. معیارهای گرایش مرکزی، جنبه‌ای مهم از مجموعه‌ی داده‌ها را توصیف می‌کنند که

همان «معدل» یا «وسط» داده‌هاست اما اطلاعاتی درباره‌ی این مشخصه‌ی اساسی دیگر، به ما نمی‌دهند. بنابراین، معرفی معیارهایی که پراکندگی در داده‌ها، یا تفرق آنها را نشان دهد، ضروری است.

مثال ۳-۱۱. نمرات دو دانشجوی رشته‌ی علوم کامپیوتر، در جدول زیر درج شده است.

ریاضی عمومی	فیزیک	اصول کامپیوتر	آمار و احتمال	
۱۳	۱۳	۱۲	۱۴	دانشجو A
۱۸	۱۳	۸	۱۳	دانشجو B

ملاحظه می‌کنیم که میانگین دو دانشجو در این چهار درس ۱۳ است. درحالی که دانشجوی A در همه‌ی دروس نمره‌ی قبولی دارد، ولی دانشجوی B در درس ریاضی عمومی قوی و در درس اصول کامپیوتر ضعیف است. بنابراین نیاز به وجود معیاری که اختلاف موجود بین این دو مجموعه داده‌های متفاوت ولی با میانگین یکسان را نشان دهد، آشکار می‌شود. توجه کنید که علاوه بر میانگین، مد و میانه‌ی نمرات هر دانشجو ۱۳ است.

مثال ۳-۱۱ نشان می‌دهد که مفهوم پراکندگی نقشی مهم در تحلیل آماری دارد. در بخش‌های ۳-۲-۱ و ۳-۲-۲ اندازه‌هایی برای پراکندگی که شامل کمیت‌های دامنه، انحراف معیار، انحراف میانگین و واریانس هست، ارائه می‌دهیم.

۳-۲-۱ دامنه

ساده‌ترین معیار سنجش پراکندگی، دامنه است که با پیدا کردن تفاضل کوچکترین مقدار داده‌ها از بزرگترین مقدار داده‌ها به دست می‌آید.

$$R = \max_{i=1,2,\dots,n} x_i - \min_{i=1,2,\dots,n} x_i \quad (3-13)$$

در مثال ۳-۱۱، دامنه‌ی نمره‌های دانشجوی A، $R_A = 14 - 12 = 2$ و دامنه‌ی نمره‌های دانشجوی B، $R_B = 18 - 8 = 10$ است. به طوری که ملاحظه می‌کنیم «پراکندگی» نمرات دانشجوی B بیشتر از دانشجوی A، است. با آنکه دامنه به سادگی

به دست می آید، ولی نقطه‌ی ضعف عمده‌ای دارد که تنها داده‌های فرین، یعنی بیشترین و کمترین مقدار داده‌ها را به حساب می آورد و کاری به مقادیر بینابینی داده‌ها ندارد. همچنین در برخی مواقع وقتی اندازه‌ی نمونه بزرگ باشد، مقدار دامنه افزایش می یابد. با این حال، وقتی که تعداد داده‌ها کم است، دامنه، معیاری مناسب برای پراکندگی است.

۲-۲-۳ انحراف معیار

برای معرفی انحراف معیار که مفیدترین معیار سنجش پراکندگی است، می توانیم میزان دوری یا نزدیکی داده‌ها را از میانگین، به عنوان ملاکی در نظر گیریم. برای توضیح مطلب، فرض کنید که داده‌های موردنظر را با مقادیر $x_1, x_2, \dots, x_n$ و میانگین آنها را با $\bar{x}$ نشان دهیم. اگر انحراف هر داده را از میانگین، یعنی مقادیر $x_1 - \bar{x}, x_2 - \bar{x}, \dots, x_n - \bar{x}$ را حساب کنیم، به نظر می رسد که همین انحراف‌ها از میانگین می توانند به عنوان معیاری برای سنجش پراکندگی داده‌ها به کار روند. به عبارت دیگر هر قدر که تفاوت این انحراف‌ها از هم بیشتر باشد، داده‌ها پراکنده ترند و هر قدر که این انحراف‌ها تفاوت کمتری باهم داشته باشند، داده‌ها متمرکزترند. چون مقایسه‌ی این n انحراف با هم کار دشواری است، برای سهولت در وهله‌ی اول سعی می کنیم میانگین این n انحراف را به دست آوریم، یعنی:

$$\frac{(x_1 - \bar{x}) + (x_2 - \bar{x}) + \dots + (x_n - \bar{x})}{n}$$

را محاسبه می کنیم. به سادگی ملاحظه می شود که بنابر تعریف میانگین، صورت این عبارت برابر صفر است و لذا میانگین انحراف‌ها از میانگین صفر است. پس می توان نتیجه گرفت: داده‌ها مجموعاً همان قدر که در جهت منفی از $\bar{x}$ انحراف دارند، مجموعاً همان قدر هم در جهت مثبت از $\bar{x}$ انحراف دارند و بنابراین مجموع انحراف‌ها صفر است.

$$\sum_{i=1}^n (x_i - \bar{x}) = 0$$

برای رفع این مشکل، باید توجه کرد که علاقه‌ی ما به‌میزان دوری داده‌ها از میانگین است و نه جهت دوری داده‌ها از آن. به‌عبارت دیگر می‌توانیم علامت جبری انحراف‌ها از میانگین را نادیده بگیریم و مثلاً

$$D_{\bar{x}} = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + \dots + |x_n - \bar{x}|}{n} \quad (۳-۱۴)$$

را محاسبه کنیم. این میانگین قدرمطلق انحراف‌ها که به **انحراف از میانگین (میانگین انحراف از میانگین)** موسوم است، کاربردهایی در آمار دارد، اما چون قدرمطلق در این تعریف مشکلاتی را ایجاد می‌کند از در نظر گرفتن آن صرف‌نظر کرده و سعی می‌کنیم مشکل علامت‌ها را در مقادیر انحراف از میانگین به‌گونه‌ای دیگر حل کنیم. برای این منظور، مقادیر انحراف از میانگین را به توان دوم می‌رسانیم، یعنی مقادیر

$$(x_1 - \bar{x})^2, (x_2 - \bar{x})^2, \dots, (x_n - \bar{x})^2$$

را محاسبه می‌کنیم. حال برای محاسبه‌ی میانگین مجذور انحراف‌ها از میانگین، باید مجموع آنها را محاسبه و بر n ، یعنی تعداد داده‌ها، تقسیم کنیم. به‌دلایلی که در اینجا نیازی به ذکر نیست، مجموع را به‌جای تقسیم بر n بر $n-1$ تقسیم می‌کنیم (تنها به ذکر این نکته بسنده می‌کنیم که با این کار برآورد دقیق‌تری از واریانس جامعه به دست می‌آید). مقدار حاصل، یعنی:

$$\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n-1}$$

را **واریانس نمونه‌ای** یا به اختصار **واریانس** می‌نامیم و آن را با s^2 نشان می‌دهیم. پس با استفاده از نماد مجموع‌یابی Σ داریم:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2. \quad (۳-۱۵)$$

چون واریانس برحسب مجذور واحدی که داده‌ها برحسب آن بیان شده‌اند، محاسبه می‌شود، برای آنکه از واحد اصلی استفاده کنیم، جذر مثبت واریانس را محاسبه و آن را با s نشان می‌دهیم و **انحراف معیار نمونه‌ای** یا صرفاً **انحراف معیار** می‌نامیم. پس

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}.$$

اگر x_i ها مقادیر جامعه‌ی آماری N عضوی باشند، آنگاه واریانس جامعه را با نماد σ^2 نمایش می‌دهیم.

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2.$$

مثال ۳-۱۲. در مثال ۳-۱۱، ملاحظه کردیم که میانگین نمرات هر دو دانشجو ۱۳ است. مقدار انحراف معیار برای دانشجوی A برابر است با:

$$s_1 = \sqrt{\frac{1}{4-1} \{ (14-13)^2 + (12-13)^2 + (13-13)^2 + (13-13)^2 \}} = \sqrt{\frac{2}{3}} = 0,816$$

همچنین انحراف معیار برای دانشجو B برابر است با:

$$s_2 = \sqrt{\frac{1}{4-1} \{ (13-13)^2 + (8-13)^2 + (13-13)^2 + (18-13)^2 \}} = \sqrt{\frac{50}{3}} = 4,082$$

ملاحظه می‌شود که انحراف معیار نمرات دانشجوی B به مراتب بیشتر از انحراف معیار دانشجوی A است و لذا این معیار پراکنده‌تر بودن نمرات دانشجوی B را در مقایسه با نمرات دانشجوی A به خوبی نشان می‌دهد.

از لحاظ محاسباتی، پیدا کردن انحراف هر داده از میانگین داده‌ها و مجموع مجذورهای انحراف‌ها برای محاسبه‌ی انحراف معیار شاید کمی مشکل باشد. به جای آن می‌توان تعریف میانگین را به کار برد و با اندکی اعمال جبری، فرمول انحراف معیار را به صورت معادل فرمول (۳-۱۶) در آورد.

$$s = \sqrt{\frac{1}{n(n-1)} \left[n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \right]}. \quad (3-16)$$

مثال ۳-۱۳. طول مدت اجرای پنج محاسبه‌ی آماری متفاوت برحسب یک دهم ثانیه، برای نرم‌افزاری به شرح زیر است.

$$0,98 \quad 0,94 \quad 1,96 \quad 14,8 \quad 2,05$$

میانگین و انحراف معیار این مجموعه‌ی داده‌ها را محاسبه کنید.

حل. بنابر تعریف، داریم

$$\bar{x} = \frac{0,98 + 0,94 + 1,96 + 14,8 + 2,05}{5} = 4,82$$

و چون

$$\sum_{i=1}^5 x_i^2 = (0.98)^2 + (0.94)^2 + (1.96)^2 + (1.48)^2 + (2.05)^2 = 12.785$$

و

$$\left(\sum_{i=1}^5 x_i\right)^2 = (0.98 + 0.94 + 1.96 + 1.48 + 2.05)^2 = 54.9081$$

پس

$$s = \sqrt{\frac{5 \times (12.785) - (54.9081)}{5 \times 4}} = 0.524.$$

اگر داده‌ها در جدول فراوانی رده‌بندی شده باشند، در این صورت با قراردادن مرکز هر رده به جای داده‌های واقع در آن رده و در نظر گرفتن فراوانی هر رده، فرمول انحراف معیار برای داده‌های رده‌بندی شده به صورت فرمول (۳-۱۷) است.

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^k (x_i - \bar{x})^2 f_i} \quad (3-17)$$

یا اگر بخواهیم محاسبات را ساده‌تر کنیم، می‌توانیم معادل فرمول (۳-۱۸) را به کار ببریم.

$$s = \sqrt{\frac{1}{n(n-1)} \left[n \left(\sum_{i=1}^k x_i^2 f_i \right) - \left(\sum_{i=1}^k x_i f_i \right)^2 \right]} \quad (3-18)$$

مثال ۳-۱۴. می‌خواهیم واریانس و انحراف معیار داده‌های مربوط به میزان اکسید سولفور منتشره در هوا را که در مثال ۲-۳ مطرح شده است، محاسبه کنیم. برای این منظور در جدول فراوانی داده‌های مذکور، ستون جدیدی اضافه می‌کنیم و در آن مقادیر $x_i^2 f_i$ را درج می‌کنیم، جدول فراوانی به صورت جدول ۳-۲ خواهد بود.

جدول ۳-۲. جدول فراوانی با دو ستون اضافی برای محاسبه‌ی میانگین و واریانس

L_U	x	f	xf	x ² f
۴,۹۵-۸,۹۵	۶,۹۵	۳	۲۰,۸۵	۱۴۴,۹۰۷۵
۸,۹۵-۱۲,۹۵	۱۰,۹۵	۱۰	۱۰۹,۵۰	۱۱۹۹,۰۲۵۰
۱۲,۹۵-۱۶,۹۵	۱۴,۹۵	۱۴	۲۰۹,۳۰	۳۱۲۹,۰۳۵۰
۱۶,۹۵-۲۰,۹۵	۱۸,۹۵	۲۵	۴۷۳,۷۵	۸۹۷۷,۵۶۲۵
۲۰,۹۵-۲۴,۹۵	۲۲,۹۵	۱۷	۳۹۰,۱۵	۸۹۵۳,۹۴۲۵
۲۴,۹۵-۲۸,۹۵	۲۶,۹۵	۹	۲۴۲,۵۵	۶۵۳۶,۷۲۲۵
۲۸,۹۵-۳۲,۹۵	۳۰,۹۵	۲	۶۱,۹۰	۱۹۱۵,۸۰۵۰
مجموع		۸۰	۱۵۰۸	۳۰۸۵۷,۰۰۰۰

بنابراین:

$$s = \sqrt{\frac{80 \times (30857) - (1508)^2}{80 \times 79}} = 5,55.$$

به‌طوری که از مقادیر واقع در آخرین ستون جدول فراوانی ۳-۲ آشکار می‌شود، محاسبات مربوط به انحراف معیار، زمانی که نماینده‌های داده‌ها ارقام اعشاری و فراوانی‌ها نسبتاً بزرگ باشند، کار خسته‌کننده‌ای است. برای ساده کردن محاسبات، می‌توان از روش کدگذاری استفاده کرد، به‌شرط اینکه طول رده‌ها برابر باشند (که در مثال ۳-۱۴ چنین است). در روش کدگذاری، نماینده‌ی یکی از رده‌ها (ترجیحاً رده‌ی وسطی یا یکی از رده‌های واقع در اواسط جدول) را صفر اختیار می‌کنیم. سپس نماینده‌های رده‌های بالای این رده را اعداد صحیح منفی، یعنی -۱، -۲، ... و نماینده‌های رده‌های واقع در زیر این رده را اعداد صحیح مثبت، یعنی ۱، ۲، ... اختیار می‌کنیم. اگر این نماینده‌های جدید رده‌ها را با $u_1, u_2, \dots, u_n$ نشان دهیم، فرمول میانگین به صورت:

$$\bar{x} = x_0 + \left(\frac{1}{n} \sum_{i=1}^k u_i f_i \right) \times c \quad (3-19)$$

درمی‌آید، که در آن x_0 نماینده‌ی رده‌ای است که در کدگذاری نماینده‌ی جدید صفر را برای آن در نظر گرفته‌ایم و c طول مشترک رده‌هاست. از آنچه که شرح داده شد، می‌توان فرمول انحراف معیار را به صورت (۲۰-۳) نمایش داد.

$$s = c \sqrt{\frac{n \left(\sum_{i=1}^k u_i^2 f_i \right) - \left(\sum_{i=1}^k u_i f_i \right)^2}{n(n-1)}} \quad (20-3)$$

نحوه‌ی محاسبه‌ی میانگین و انحراف معیار را به کمک روش کدگذاری، در مثال زیر تشریح می‌کنیم.

مثال ۳-۱۵. مطابق داده‌های مثال ۳-۱۴ میانگین و انحراف معیار را به روش کدگذاری به دست آورید.

حل. چون تعداد رده‌ها هفت است، نماینده‌ی جدید رده‌ی چهارم را صفر می‌گیریم و به ترتیبی که در بالا گفتیم نماینده‌های جدید رده‌های دیگر را به دست می‌آوریم. حال جدول ۳-۳ را تشکیل می‌دهیم.

جدول ۳-۳. جدول فراوانی با اعمال کدگذاری برای مثال ۳-۱۵

m_i	u_i	f_i	$f_i u_i$	$u_i^2 f_i$
۶,۹۵	-۳	۳	-۹	۲۷
۱۰,۹۵	-۲	۱۰	-۲۰	۴۰
۱۴,۹۵	-۱	۱۴	-۱۴	۱۴
۱۸,۹۵	۰	۲۵	۰	۰
۲۲,۹۵	۱	۱۷	۱۷	۱۷
۲۶,۹۵	۲	۹	۱۸	۳۶
۳۰,۹۵	۳	۲	۶	۱۸
		$\sum f_i = ۸۰$	$\sum f_i u_i = -۲$	$\sum u_i^2 f_i = ۱۵۲$

با قراردادن $c=۴$ ، $x_0=۱۸,۹۵$ ، $n=۸۰$ ، $\sum_{i=1}^v u_i f_i = -۲$ در فرمول (۳-۱۴) میانگین به روش کدگذاری داریم:

$$\bar{x} = x_0 + \frac{\sum_{i=1}^k u_i f_i}{n} \cdot c = 18,95 + \frac{-2}{80} \times 4 = 18,85.$$

همچنین با قراردادن همین مقادیر همراه با مقدار $\sum_{i=1}^v u_i^* f_i = 152$ در فرمول (۳-۱۵) انحراف معیار به روش کدگذاری خواهیم داشت:

$$s = 4 \sqrt{\frac{80(152) - (-2)^2}{80 \times 79}} = 0,55$$

به طوری که ملاحظه می‌کنیم این مقادیر، مطابق انتظار، همان مقادیری هستند که از داده‌های اصلی به دست آمده‌اند.

۳-۲-۳ فرمول‌های دیگر برای محاسبه‌ی پراکندگی

می‌توان در فرمول (۳-۱۴) محاسبه‌ی انحراف از میانگین فرم کلی‌تری، به نام میانگین انحراف از a (عدد حقیقی)، در نظر گرفت.

$$D_a = \frac{1}{n} \sum_{i=1}^n |x_i - a| \quad (۳-۲۱)$$

به حالت‌های خاص $a = Mo$ و $a = Md$ به ترتیب D_{Mo} و D_{Md} ، «میانگین انحرافات از میانه» و «میانگین انحرافات از نما» می‌گوییم. اما ویژگی خاص که برای

میانه در معیار D_a رخ می‌دهد، این است که کمترین مقدار $\sum_{i=1}^n |x_i - a|$ به ازای $a = Md$ به دست می‌آید یعنی:

$$\min_{a \in R} \sum_{i=1}^n |x_i - a| = \sum_{i=1}^n |x_i - Md| \quad (۳-۲۲)$$

همچنین فرمول کلی‌تر دیگری برای اندازه‌ی پراکندگی، که حالت خاص آن به عنوان «واریانس» یاد می‌شود، به صورت (۳-۲۳) است.

$$MS_a = \frac{1}{n-1} \sum_{i=1}^n (x_i - a)^2, \quad (۳-۲۳)$$

که حداقل مقدار ممکنه عبارت (۳-۲۳) به ازای میانگین x_i ها برقرار است. به MS_a «میانگین مربعات انحراف از a » می‌گوییم.

$$\min_{a \in R} MS_a = MS_{\bar{x}} = S^2. \quad (۳-۲۴)$$

۳-۳ کاربردهای از معیارهای پراکندگی

برخی کاربردهای انحراف معیار، شامل ضریب تغییر، نمره‌ی استاندارد، قضیه چبیشف را در ادامه شرح می‌دهیم.

۳-۳-۱ ضریب تغییر

روش کدگذاری حالت خاصی از تبدیل متغیرها است. اگر مشاهدات را در عددی ثابت ضرب و با عددی ثابت جمع کنیم، مشاهدات جدیدی به دست می‌آیند.

$$y_i = ax_i + b \quad i=1,2,\dots,n$$

از روی خواص میانگین و واریانس، تأثیر این تبدیل بر روی آنها به شکل زیر می‌شود:

$$\bar{y} = a\bar{x} + b, \quad s_y^2 = a^2 s_x^2.$$

معیارهای میانگین و انحراف معیار دارای واحد سنجش، از نوع واحد سنجش داده‌هاست. برای مثال ممکن است وزن اشیاء مشخصی دارای انحراف معیار ۰/۱ اونس یا ۲۸۳۵ میلی‌گرم باشند که هر دو یکسان هستند، ولی هیچ‌کدام از آنها به راحتی مشخص نمی‌کنند که با یک پراکندگی بزرگ یا یک پراکندگی کوچک سروکار داریم. در این گونه موارد آنچه که مورد نیاز است ارائه‌ی یک مقیاس نسبی پراکندگی است. یک مقیاس نسبی پراکندگی عبارت است از ضریب تغییر که از تقسیم انحراف معیار بر میانگین حاصل می‌گردد که معمولاً در عدد ۱۰۰ ضرب شده و به صورت درصد بیان و با حرف V نشان داده می‌شود.

$$V = \frac{s}{\bar{x}} \times 100 \quad (۳-۲۵)$$

برای اطلاعات مثال ۳-۱۵، مقدار ضریب تغییر برابر است با:

$$V = \frac{5/55}{18/85} \times 100 = 29/4.$$

که این کمیت، نمایانگر میزان تغییرات در یک واحد اندازه‌گیری است.

۳-۲-۳ نمره‌ی استاندارد

فرض کنید $x_1, x_2, \dots, x_n$ یک سری داده n تایی با میانگین $\bar{x}$ و انحراف معیار s باشند. با تبدیل

$$z_i = \frac{x_i - \bar{x}}{s}, \quad i = 1, 2, \dots, n \quad (3-26)$$

داده‌های جدید $z_1, z_2, \dots, z_n$ به نام نمره‌های استاندارد، به دست می‌آیند. این داده‌ها دارای ویژگی میانگین صفر و واریانس یک، $\bar{z} = 0$ و $s_z^2 = 1$ هستند. هرگاه بخواهیم دو مشاهده از دو نمونه‌ی مختلف را با هم مقایسه کنیم، استانداردشده‌ی آنها کار را آسان می‌کند.

مثال ۳-۱۶. میانگین و انحراف معیار نمرات پایان‌ترم درس زبان خارجه به ترتیب ۶۰ و ۱۶ و میانگین و انحراف معیار نمرات پایان‌ترم درس ریاضی عمومی به ترتیب ۵۸ و ۱۰ است. اگر دانشجویی هر دو درس را امتحان داده باشد و نمره‌ی زبان خارجه ۷۲ و نمره‌ی ریاضی عمومی ۶۸ باشد، وضعیت تحصیلی‌اش در کدام درس بهتر است؟

$$z_1 = \frac{72 - 60}{16} = 0.75 \text{ نمره‌ی استاندارد زبان خارجه: } 0.75$$

$$z_2 = \frac{68 - 58}{10} = 1 \text{ نمره‌ی استاندارد ریاضی عمومی: } 1$$

ملاحظه می‌شود که اگرچه نمره‌ی درس زبان خارجه از نمره‌ی درس ریاضی عمومی او بیشتر است، ولی وضعیت وی در کلاس ریاضی عمومی بهتر از کلاس زبان خارجه است. زیرا نمره‌ی استاندارد در درس ریاضی عمومی به اندازه‌ی یک انحراف معیار بالای میانگین، قرار دارد در حالی که در درس زبان خارجه به اندازه‌ی ۰.۷۵ انحراف معیار در بالای میانگین، قرار دارد.

۳-۳-۳ قضیه‌ی چیشف

چنانچه پراکندگی دو نمونه از یک جمعیت را با یکدیگر مقایسه کنیم، می‌گوییم نمونه‌ای که دارای انحراف معیار بزرگتری است، پراکندگی‌اش بیشتر است. همچنین عنوان کردیم که هرچه داده‌ها به میانگین نزدیکتر باشند و یا پیرامون میانگین تجمع

کرده باشند، انحراف معیار به صفر نزدیک است و چنانچه داده‌ها از میانگین دور باشند، آنگاه انحراف معیار بزرگ است. این موضوع به گونه‌ای علمی توسط ریاضی‌دان روسی به نام چبیشف با عنوان قضیه‌ی چبیشف بیان شده است.

قضیه‌ی ۱-۳ (قضیه‌ی چبیشف). برای هر مجموعه از داده‌ها و هر عدد ثابت $k \geq 1$ ، حداقل $\left(1 - \frac{1}{k^2}\right) \times 100$ درصد داده‌ها باید به فاصله‌ی k برابر انحراف معیار در هر دو طرف میانگین قرار گیرند.

قضیه‌ی چبیشف برای حالت‌های خاص k ، چنین بیان می‌شود:

الف. $k = 1$ ، ممکن است تعداد خیلی کم داده‌ها به بازه‌ی $(\bar{x} - s, \bar{x} + s)$ تعلق داشته باشند.

ب. $k = 2$ ، حداقل $\frac{3}{4} \times 100 = 75$ درصد، داده‌ها به بازه‌ی $(\bar{x} - 2s, \bar{x} + 2s)$ تعلق دارند.

ج. $k = 3$ ، حداقل $\frac{8}{9} \times 100 = 88,9$ درصد، داده‌ها به بازه‌ی $(\bar{x} - 3s, \bar{x} + 3s)$ تعلق دارند.

مثال ۱۷-۳. در دستگاه اتوماتیک بسته‌بندی زعفران به گونه‌ای تنظیم شده است که درون هر بسته به‌طور متوسط ۱۶ گرم زعفران و با انحراف معیار ۰,۰۲ گرم، زعفران قرار می‌دهد. حداقل درصد بسته‌های تولیدی‌اش که دارای وزنی بین ۱۵,۹۵ تا ۱۶,۰۵ گرم باشند، ۸۴ درصد است (توجه دارید که در این مثال $k = 2,5$ در نظر گرفته شده است).

قضیه‌ی چبیشف برای هر نوع داده‌ای قابل کاربرد است، اما چنین بیان می‌شود که حداقل چند درصد داده‌ها به فاصله‌ی معینی تعلق دارند و مقدار دقیق را تعیین نمی‌کند. برای بیشتر داده‌ها، درصد واقعی داده‌هایی که به فاصله‌ی معینی تعلق می‌گیرند، بیش از آنچه که قضیه‌ی چبیشف بیان می‌کند، می‌باشد.

برای بیشتر داده‌ها راه‌کار دقیق‌تری از قضیه‌ی چبیشف با عنوان دستور تجربی بیان شده است. در حقیقت یک نتیجه‌ی نظری است که براساس توزیع نرمال که در مباحث بعدی احتمال مطرح می‌شود (به فصل چهارم مراجعه شود).

دستور تجربی: برای داده‌هایی که منحنی فراوانی آنها زنگی شکل است، داریم:

الف. تقریباً ۶۸ درصد داده‌ها به بازه‌ی $(\bar{x} - s, \bar{x} + s)$ تعلق دارند.

ب. تقریباً ۹۵ درصد داده‌ها به بازه‌ی $(\bar{x} - 2s, \bar{x} + 2s)$ تعلق دارند.

ج. تقریباً ۹۹٫۷ درصد داده‌ها به بازه‌ی $(\bar{x} - 3s, \bar{x} + 3s)$ تعلق دارند.

مثال ۳-۱۸. با اطلاعات مثال ۲-۳ و شکل ۲-۱ که تقارن تقریبی در نمودار به چشم می‌خورد، می‌توان وضعیت توزیع مشاهدات را در بازه‌های مطرح شده برآورد کرد.

الف. تقریباً ۶۸ درصد داده‌ها به بازه‌ی

$$(18,85 - 0,55, 18,85 + 0,55) = (13,3, 24,4)$$

تعلق دارند.

ب. تقریباً ۹۵ درصد داده‌ها به بازه‌ی

$$(18,85 - 2 \times 0,55, 18,85 + 2 \times 0,55) = (7,75, 29,95) = (0, 29,95)$$

تعلق دارند.

ج. تقریباً ۹۹٫۷ درصد (تقریباً کل) داده‌ها به بازه‌ی

$$(18,85 - 3 \times 0,55, 18,85 + 3 \times 0,55) = (2,2, 35,0)$$

تعلق دارند.

۳-۴ چندک‌ها و دیگر معیارهای پراکندگی

یادآوری می‌کنیم که میانه‌ی مجموعه‌ای از داده‌ها، مقداری بود که نصف داده‌ها در یک طرف و نصف دیگر در طرف دیگر آن قرار داشتند. به همین قیاس می‌توان چندک‌ها را تعریف کرد که عبارت از مقادیری هستند که بخش معینی از داده‌ها در یک طرف آن و بخش باقی‌مانده در طرف دیگر آن قرار دارند. به بیانی دقیق‌تر:

عدد Q_p را که در آن $0 < p < 1$ ، «چندک مرتبه‌ی p » می‌نامند، هرگاه $100p$ درصد داده‌ها کوچکتر از آن باشند.

از جمله‌ی چندک‌های مهم می‌توان از چارک‌ها $(p = 0,25, 0,5, 0,75)$ ، دهک‌ها

$(p = 0,1, 0,2, \dots, 0,9)$ ، و صدک‌ها $(p = 0,01, 0,02, \dots, 0,99)$ نام برد. این مفاهیم را

ابتدا برای داده‌های رده‌بندی شده و سپس برای داده‌های کمی-گسسته تعریف می‌کنیم.

چارک اول نقطه‌ای بر روی محور طول‌هاست که در آن نقطه، مساحت زیر بافت‌نگار به نسبت‌های ۱ و ۳ تقسیم می‌شود، یعنی $\frac{1}{4}$ مساحت زیر بافت‌نگار در سمت چپ این نقطه و $\frac{3}{4}$ سطح باقی‌مانده در طرف دیگر این نقطه قرار دارد. چارک دوم همان میانه است و چارک سوم نقطه‌ای است که $\frac{3}{4}$ مساحت زیر بافت‌نگار در سمت چپ آن و $\frac{1}{4}$ در سمت راست آن قرار دارد. چارک اول را با Q_1 و چارک دوم (یا همان میانه) را با Q_2 و چارک سوم را با Q_3 نشان می‌دهیم.

محاسبه‌ی Q_1 برای داده‌های پیوسته (که از مقیاس‌های فاصله و نسبت به‌دست آمده‌اند) از فرمول (۳-۲۷) استفاده می‌شود.

$$Q_1 = L_k + \frac{\frac{n}{4} - F_{k-1}}{f_k} \times c \quad (3-27)$$

که در آن L_k کران پایین طبقه‌ی چارک اول، F_{k-1} فراوانی تجمعی طبقه‌ی ماقبل طبقه‌ی چارک اول، f_k فراوانی طبقه‌ی چارک اول، c طول رده‌ی چارک اول است.

درست به همین منوال می‌توانیم چارک سوم، Q_3 ، را نیز محاسبه کنیم. در این حالت با استفاده از فرمول (۳-۲۸) مقدار عددی چارک سوم به‌دست می‌آید.

$$Q_3 = L_k + \frac{\frac{3n}{4} - F_{k-1}}{f_k} \times c \quad (3-28)$$

که در آن L_k کران پایین طبقه‌ی چارک سوم، F_{k-1} فراوانی تجمعی طبقه‌ی ماقبل طبقه‌ی چارک سوم، f_k فراوانی طبقه‌ی چارک سوم، c طول رده‌ی چارک سوم است.

به‌همین ترتیب می‌توان دهک‌های اول، دوم، ...، نهم را تعریف کرد. این دهک‌ها به ترتیب با D_1 ، D_2 ، ...، D_9 نشان داده می‌شوند. مثلاً دهک اول، از فرمول (۳-۲۹) محاسبه می‌شود.

$$D_1 = L_k + \frac{\frac{n}{10} - F_{k-1}}{f_k} \times c \quad (3-29)$$

که در آن L_k کران پایین طبقه‌ی دهک اول، F_{k-1} فراوانی تجمعی طبقه‌ی ماقبل طبقه‌ی دهک اول، f_k فراوانی طبقه‌ی دهک اول، c طول رده‌ی دهک اول است. صدک‌ها عیناً به‌نحو مشابه تعریف می‌شوند که از ذکر جزئیات مربوط به آنها خودداری می‌کنیم.

یکی دیگر از معیارهای پراکندگی، **برد میان‌چارکی** است که به‌صورت $QR = Q_3 - Q_1$ تعریف می‌شود. **برد میان‌چارکی** طول بازه‌ای است که پنجاه درصد داده‌ها در آن قرار دارند. برخی پژوهشگران از **نیم برد میان‌چارکی** استفاده می‌کنند که به‌صورت:

$$IQR = \frac{Q_3 - Q_1}{2}$$

تعریف می‌شود، و **انحراف چارکی** نیز نامیده می‌شود. یک اندازه‌ی پراکندگی نسبی، که ضریب تغییر چارکی نام دارد به‌صورت نسبت نیم برد میان‌چارکی بر میانگین Q_1 و Q_3 ضرب در ۱۰۰، یعنی به‌صورت:

$$VQ = \frac{Q_3 - Q_1}{Q_3 + Q_1} \times 100$$

تعریف می‌شود.

می‌توان سایر دهک‌ها را به‌همین ترتیب محاسبه و مثلاً $D_4 - D_1$ یا $D_8 - D_7$ ، و ... را به‌عنوان معیارهای دیگری برای سنجش پراکندگی به‌کار برد.

مثال ۳-۱۹. اطلاعات مثال ۲-۳، مربوط به میزان اکسید سولفور، را در نظر بگیرید. مقادیر چارک اول، سوم، ضریب تغییر چارکی، دهک اول و دهک نهم را به‌دست آورید.

حل. چون $n = 80$ ، پس $\frac{n}{4} = 20$. با توجه به اینکه

$$3 + 10 < 20 < 3 + 10 + 14$$

پس چارک اول در رده‌ی سوم قرار دارد که کران پایین آن $L_1 = 12,95$ و فراوانی آن $f_3 = 14$ است. چون $\frac{n}{4} - F_{k-1} = 20 - (3 + 10) = 7$ ، بنابراین مطابق فرمول (۳-۲۷) داریم:

$$Q_1 = 12,95 + \frac{7}{14} \times 4 = 14,95.$$

همچنین، چون $\frac{3n}{4} = 60$ ، و

$$3 + 10 + 14 + 25 < 60 \leq 3 + 10 + 14 + 25 + 17$$

پس

$$Q_3 = 20,95 + \frac{1}{17} \times 4 = 22,83.$$

با توجه به تعریف، برد میان‌چارکی

$$Q_3 - Q_1 = 22,83 - 14,95 = 7,88$$

و لذا نیم برد میان‌چارکی $IQR = \frac{Q_3 - Q_1}{2} = 3,94$ و ضریب تغییر چارکی

$$VQ = \frac{Q_3 - Q_1}{Q_3 + Q_1} \times 100 = 20,86.$$

چون $\frac{n}{10} = 8$ و $3 < 8 < 3 + 10$ پس دهک اول در رده‌ی دوم با کران پایین $8/95$ و فراوانی ۱۰ قرار دارد. چون $h = 8 - 3 = 5$ ، پس

$$D_1 = 8,95 + \frac{5}{10} \times 4 = 10,95.$$

به همین قیاس، چون $\frac{9n}{10} = 72$ و

$$3 + 10 + 14 + 25 + 17 < 72 \leq 3 + 10 + 14 + 25 + 17 + 9$$

لذا دهک نهم در رده‌ی ششم با کران پایین $24/95$ قرار دارد. در نتیجه

$$D_9 = 24,95 + \frac{72 - (3 + 10 + 14 + 25 + 17)}{9} \times 4 = 24,95 + \frac{4}{3} = 26,28$$

همان‌طور که پیش‌تر گفتیم، می‌توان مثلاً از $D_9 - D_1$ نیز به عنوان معیاری برای سنجش پراکندگی استفاده کرد. در این مثال $D_9 - D_1 = 15,33$. توجه کنید که ۸۰ درصد داده‌ها در بازه‌ی D_1 تا D_9 قرار دارد.

مطابق با آنچه که گذشت، فرمول کلی برای محاسبه‌ی چندک مرتبه‌ی p در جداول آماری با داده‌های پیوسته، عبارت است از:

$$Q_p = L_k + \frac{pn - F_{k-1}}{f_k} \times c, \quad (3-30)$$

که در آن L_k کران پایین طبقه‌ی چندک مرتبه‌ی p ، F_{k-1} فراوانی تجمعی طبقه‌ی ماقبل طبقه‌ی چندک مرتبه‌ی p ، f_k فراوانی طبقه‌ی چندک مرتبه‌ی p ، c طول رده‌ی چندک مرتبه‌ی p است.

برای محاسبه‌ی چندک مرتبه‌ی p ، در داده‌های گسسته (همانگونه که در محاسبه‌ی میانه ملاحظه شد)، ابتدا n داده را به صورت غیرنزولی مرتب می‌کنیم،

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

سپس جزء صحیح مقدار $(n+1)p$ را با r و قسمت اعشاری آن را w در نظر می‌گیریم. واضح است:

$$r = [(n+1)p], \quad w = (n+1)p - r, \quad 0 \leq w < 1$$

اگر $w=0$ ، آنگاه $Q_p = x_{(r)}$ ، در غیر این صورت

$$Q_p = (1-w)x_{(r)} + wx_{(r+1)}. \quad (3-31)$$

به‌سادگی می‌توان نشان داد که فرمول (3-31) از فرمول (3-30) قابل محاسبه است.

مثال ۳-۲۰. برای داده‌های ۳، ۴، ۵، ۶، ۷، ۸، ۹، ۱۰ دهک هفتم را به‌دست آورید.

حل. داریم $n=9$ ، $p=0.7$ ، $(n+1)p = 10 \times 0.7 = 7$ ، $r=7$ و $w=0$ پس

$$Q_{0.7} = x_{(7)} = 9$$

مثال ۳-۲۱. برای داده‌های ۰، ۱، ۲، ۳، ۴، ۵، ۶، ۷، ۸، ۱۲ صدک شصتم را به‌دست آورید.

حل. داریم $n=10$ ، $p=0.6$ ، $(n+1)p = 11 \times 0.6 = 6.6$ ، $r=6$ و $w=0.6$ پس

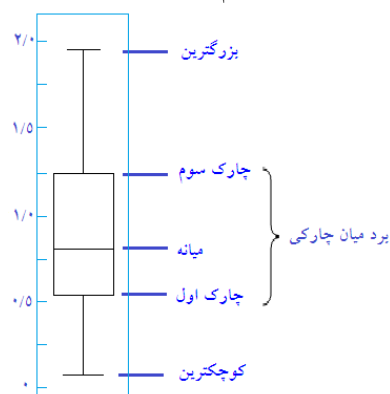
$$Q_{0.6} = (1-0.6)x_{(6)} + 0.6x_{(7)} = 0.4 \times 5 + 0.6 \times 7 = 6.2.$$

۳-۵ نمودار جعبه‌ای

نمودار جعبه‌ای برای نمایش فواصل بین مقادیر چارک اول، چارک سوم، میانه، و

میانگین بسیار سودمند است. در این نمودار، خارج از یک سر جعبه چارک اول و خارج از سر دیگر چارک سوم قرار دارد، بزرگترین و کوچکترین مشاهده نمایش داده می‌شود. این فواصل در نمودار جعبه‌ای، می‌تواند حاکی از وجود نقاط پرت، و همچنین عدم تقارن توزیع داده‌ها، تفسیر شود. همچنین نمودار جعبه‌ای برای توصیف و مقایسه‌ی اطلاعات چند جامعه نیز به‌کار می‌رود.

فرم عمومی نمودار جعبه‌ای در شکل ۳-۱ آمده است. این نمودار از اطلاعاتی مانند کوچکترین، چارک اول، میانه، چارک سوم، بزرگترین مشاهدات ساخته شده است.



شکل ۳-۱. فرم عمومی نمودار جعبه‌ای

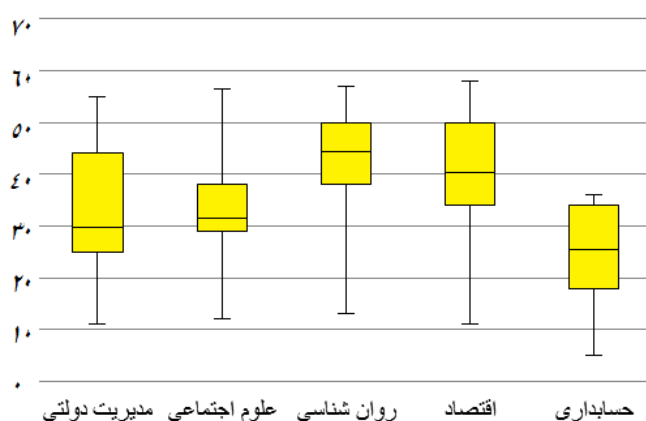
مثال ۳-۲۲ اطلاعات مقایسه‌ای بین چند جامعه

در یک بررسی از پنج رشته‌ی مدیریت دولتی، اقتصاد، علوم اجتماعی، روان‌شناسی، و حسابداری در یکی از مراکز دانشگاه پیام‌نور؛ تعداد روزهای سپری شده بعد از انتخاب واحد تا روز پرداخت شهریه، ثبت شده است. نمودار مقایسه‌ای پنج رشته، در نمودار ۳-۲ آمده است.

جدول ۳-۴. آماره‌های توصیفی پنج رشته‌ی مثال ۳-۲۲

رشته‌ی تحصیلی	max	Q_3	Md	Q_1	min
مدیریت دولتی	۵۵	۴۴	۳۰	۲۵	۱۱
علوم اجتماعی	۵۷	۳۸	۳۲	۲۹	۱۲
روان‌شناسی	۵۷	۵۰	۴۵	۳۸	۱۳
اقتصاد	۵۸	۵۰	۴۰	۳۴	۱۱

حسابداری	۳۶	۳۴	۲۵	۱۸	۵
----------	----	----	----	----	---



نمودار ۱-۳. نمودار جعبه‌ای تعداد روزهای سپری‌شده تا پرداخت شهریه

این نمودار حاکی از آن است که دانشجویان رشته‌ی روان‌شناسی، به نسبت سایر رشته‌ها، دیرتر اقدام به پرداخت شهریه دارند. این تعبیر، بیشتر باید براساس طول و موقعیت مستطیل‌ها باشد، تا اندازه‌ی طول دم‌های خارج از مستطیل‌ها.

خلاصه‌ی فصل سوم

در این فصل انواع معیارهای مرکزی و پراکندگی و نمودار جعبه‌ای معرفی شدند. به جز مد، که می‌تواند برای مشاهدات از روی یک متغیر کیفی نیز محاسبه شود، مابقی معیارها به متغیرهای کمی اختصاص دارد. محاسبه‌ی انواع میانگین‌ها و انحراف معیار و واریانس برای داده‌های جدول‌شده با استفاده از نماینده‌ی رده‌ها، به جای داده‌های اصلی انجام می‌شود. مفاهیمی مانند نمره‌ی استاندارد و ضریب تغییرات برای مقایسه‌ی دو یا چند جامعه یا چند عضو به کار می‌رود. جدول زیر فرمول‌های هریک از معیارهای ارائه‌شده را نمایش می‌دهد. توجه دارید که L_k کران پایین طبقه‌ی میانه، F_{k-1} فراوانی تجمعی طبقه‌ی ماقبل طبقه‌ی میانه، f_k فراوانی طبقه‌ی میانه، c طول رده‌ی میانه، و $0 \leq w < 1$ ، $w = (n+1)p - r$ ، $r = [(n+1)p]$ تعریف شده است.

جدول ۳ - ۵. خلاصه‌ای از معیارهای مرکزی و پراکندگی

نوع	عنوان	محاسبه از روی داده‌ها	محاسبه از روی جدول
مرکزی	میانگین حسابی	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$	$\bar{x} = \frac{1}{n} \sum_{i=1}^k f_i x_i$
	میانه	$Md = \begin{cases} x_{(\frac{n+1}{2})} & \text{فرد } n \\ \frac{1}{2}(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}) & \text{زوج } n \end{cases}$	$Md = L_k + \frac{\frac{n}{2} - F_{k-1}}{f_k} \times c$
	میانگین توافقی	$\bar{x}_H = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$	$\bar{x}_H = \frac{n}{\sum_{i=1}^k \frac{1}{x_i} f_i}$
	میانگین هندسی	$\bar{x}_G = \sqrt[n]{x_1 x_2 \cdots x_n}$	$\bar{x}_G = \exp\left(\frac{1}{n} \sum_{i=1}^k f_i \log(x_i)\right)$
	میانگین رتبه‌ی دو	$M_r = \left(\frac{1}{n} \sum_{i=1}^n x_i^r\right)^{\frac{1}{r}}$	$M_r = \left(\frac{1}{n} \sum_{i=1}^k f_i x_i^r\right)^{\frac{1}{r}}$
	چارک اول		$Q_1 = L_k + \frac{\frac{n}{4} - F_{k-1}}{f_k} \times c$
	چارک سوم		$Q_3 = L_k + \frac{\frac{3n}{4} - F_{k-1}}{f_k} \times c$
	چندک مرتبه p	$Q_p = (1-w)x_{(r)} + wx_{(r+1)}$	$Q_p = L_k + \frac{pn - F_{k-1}}{f_k} \times c$
پراکندگی	دامنه		$R = \max_{i=1,2,\dots,n} x_i - \min_{i=1,2,\dots,n} x_i$
	واریانس	$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$	$s^2 = \frac{1}{n-1} \sum_{i=1}^k (x_i - \bar{x})^2 f_i$

انحراف معیار	$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$	$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^k (x_i - \bar{x})^2 f_i}$
برد میان چارکی	$QR = Q_3 - Q_1$	
ضریب تغییر چارکی	$VQ = \frac{Q_3 - Q_1}{Q_3 + Q_1} \times 100$	

تمرین

۳-۱. برای مشاهدات زیر میانه، نما و چارک اول را محاسبه کنید.

۲ ۸ ۳ ۷ ۰ ۶ ۱ ۳ ۴ ۲ ۱ ۳ ۰ ۱ ۵ ۲ ۳ ۰ ۴

۳-۲. سه نقطه A ، B ، و C به یک فاصله از یکدیگر قرار دارند. متحرکی از نقطه‌ی A به نقطه‌ی B ، از نقطه‌ی B به نقطه‌ی C و از نقطه‌ی C به نقطه‌ی A به ترتیب با سرعتی معادل ۲۰ متر در ساعت، ۴۰ متر در ساعت و ۸۰ متر در ساعت حرکت می‌کند. سرعت متوسط این متحرک در مسیر ذکر شده را محاسبه کنید.

۳-۳. رابطه‌ی زیر را به ازای هر a ثابت کنید

$$\sum_{i=1}^n (x_i - a)^2 = \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - a)^2.$$

۳-۴. آسانسوری به گونه‌ای طراحی شده است که بیشترین بار آن ۲۰۰۰ پوند باشد. اگر در یک جابه‌جایی مسافران این آسانسور ۸ مسافر زن که میانگین وزن آنها ۱۲۳ پوند است و ۵ مرد که میانگین وزن آنها ۱۷۴ پوند است، باشند آیا آسانسور دارای بار بیش از حد مجاز است؟

۳-۵. سه نمایندگی فروش یخچال از شرکتی، در سه شهر بزرگ تعداد ۵۷۵، ۴۱۰، و ۴۲۵ یخچال را به ترتیب با متوسط قیمت ۱۲۰۰۰ تومان، ۱۱۷۰۰ تومان و ۱۲۱۰۰ تومان فروخته‌اند. متوسط قیمت یخچال‌های فروخته شده چقدر است؟

- ۳- ۶. اگر واریانس یک جمعیت برابر با ۲ و میانگین آنها برابر با ۳ و مجموع توان‌های دوم آنها برابر با ۵۵ باشد، اندازه‌ی جمعیت چقدر است؟
- ۳- ۷. اگر میانگین داده‌های $x_1, x_2, \dots, x_n$ برابر با $\bar{x}$ باشد، آنگاه میانگین داده‌های $(x_1 + 1), (x_2 + 1), \dots, (x_n + 1)$ چقدر است؟
- ۳- ۸. اگر میانگین داده‌های $x_1, x_2, \dots, x_n$ برابر با ۱۰ باشد، آنگاه میانگین داده‌های $(x_1 + 9), (x_2 + 9), \dots, (x_n + 9)$ را محاسبه کنید.
- ۳- ۹. اگر میانگین داده‌های $x_1, x_2, \dots, x_n$ برابر با ۱۵ باشد، آنگاه میانگین داده‌های $(x_1 + 8), (x_2 + 8), \dots, (x_n + 8)$ را محاسبه کنید.
- ۳- ۱۰. اگر میانگین داده‌های $x_1, x_2, \dots, x_n$ برابر با ۱۵ باشد، آنگاه میانگین داده‌های ۳۶، ۳۵، ۳۰، $x_1, x_2, \dots, x_n$ را به دست آورید.
- ۳- ۱۱. اگر انحراف معیار داده‌های x_1, x_2, x_3, x_4 برابر با صفر باشد، آنگاه میانگین داده‌های ۵۰، x_1, x_2, x_3, x_4 را به دست آورید.
- ۳- ۱۲. اگر به هریک از داده‌ها عدد ۱۵ را بیفزاییم، در مقدار انحراف میانگین چه تغییری رخ می‌دهد؟
- ۳- ۱۳. اگر انحراف معیار داده‌هایی ۱۵، $x_1, x_2, \dots, x_n$ برابر با صفر باشد، میانه‌ی داده‌های ۱۷، ۱۹، $(2x_1 + 1), \dots, (2x_n + 1)$ را محاسبه کنید.
- ۳- ۱۴. اگر انحراف معیار اعداد ۳۶، ۲۴، ۲۰، ۱۲ و ۸ برابر با $9/8$ باشد، انحراف معیار اعداد ۹، ۶، ۵، ۳ و ۲ چقدر است؟
- ۳- ۱۵. اگر میانه‌ی داده‌های $x_1, x_2, \dots, x_n$ برابر با $5/6$ باشد، آنگاه میانه‌ی $x_1, x_2, \dots, x_n, 4/5, 17/6$ چقدر است؟

خودآزمایی چهارگزینه‌ای

- ۳- ۱. اگر میانگین $x_1, x_2, \dots, x_n$ برابر با ۱۶ باشد، آنگاه کدام گزینه مقداری کمتری برای عبارت $\sum_{i=1}^n (x_i - a)^2$ ایجاد می‌کند.
- الف. صفر ب. ۱۶ ج. $16 - n$ د. $16 + n$

۳-۲. میانگین نمرات گروه یک ۱۴ نفری برابر با ۱۶ و میانگین نمرات گروه دو ۲۰

نفری برابر ۱۳ است. میانگین کل ۳۴ نفری برابر کدام گزینه است؟

الف. ۱۴/۵ ب. ۱۴/۲۵ ج. ۱۵ د. ۱۴

۳-۳. در چه صورت عبارت $\sum_{i=1}^k \frac{i}{n} f_i$ نمایانگر میانگین است؟

الف. نماینده‌ی رده‌ها اعداد $1, 2, \dots, k$ باشند.

ب. نماینده‌ی رده‌ها اعداد $\frac{1}{n}, \frac{2}{n}, \dots, \frac{k}{n}$ باشند.

ج. نماینده‌ی رده‌ها اعداد $1, 2, \dots, n$ باشند.

د. نماینده‌ی رده‌ها اعداد $\frac{1}{k}, \frac{2}{k}, \dots, \frac{k}{k}$ باشند.

۳-۴. میانه‌ی مشاهدات ۱۰۴، ۹۲، ۸۰، ۴۵، ۴۴ برابر کدام گزینه است؟

الف. ۸۰ ب. ۶۲/۵ ج. ۸۶ د. ۸۷

۳-۵. اگر فراوانی رده‌های ۲۴-۲۰، ۲۰-۱۶، ۱۶-۱۲، ۱۲-۸، ۸-۴ به ترتیب ۱۰، ۷، ۵، ۴، ۸ باشند، آنگاه میانه برابر کدام گزینه است؟

الف. ۱۶ ب. ۱۲ ج. ۱۸ د. ۲۰

۳-۶. اگر مشاهدات ۱۹، ۱۸، ۱۵، ۱۱، ۱۰، ۸، ۵، ۳، ۲، ۱ دارای میانگین ۹/۲ باشد، افزودن

کدام گزینه به مشاهدات، تغییری در میانگین مشاهدات جدید ایجاد نمی‌کند؟

الف. ۰، ۰، ۰ ب. ۰، ۰، ۰، ۰، ۰، ۰، ۰، ۰، ۰، ۰ ج. ۰، ۰، ۰، ۰، ۰، ۰، ۰، ۰، ۰، ۰ د. ۰، ۰، ۰، ۰، ۰، ۰، ۰، ۰، ۰، ۰

۳-۷. میانگین مشاهدات a^1, a^2, a^3, a^4 برابر کدام گزینه است؟

الف. a^1 ب. a^0 ج. a^2 د. a^3

۳-۸. اگر بدانیم میانگین حسابی ۲/۵ و میانگین توافقی ۱/۸ است، آنگاه کدام گزینه

برای میانگین هندسی برقرار است؟

الف. کمتر از ۱/۸ ب. بیش از ۲/۵ ج. بین دو میانگین داده شده د. برابر ۲/۱

۳-۹. اگر $\sum_{i=1}^n x_i = k$ و $\sum_{i=1}^n x_i^2 = k^2$ ، آنگاه واریانس نمونه‌ای برابر با کدام گزینه

است؟

الف. $\frac{k^2}{n}$ ب. $\frac{k^2}{8}$ ج. $\frac{k}{n}$ د. $\frac{k}{8}$

۳- ۱۰. سه عدد $x \leq y \leq z$ را طوری در نظر بگیرید که میانگین آنها y باشد. کدام گزینه واریانس یک را برای آنها برقرار می‌کند؟

الف. $z = x = y$ ج. $(x - y)^2 = 1$

ب. $(z - y)^2 = 1$ د. x و z بر روی دایره‌ای به شعاع $y\sqrt{2}$ قرار گیرند.

۳- ۱۱. اگر میانگین مشاهدات برابر ۱۴ و واریانس آن برابر ۴۹ باشد، آنگاه مقدار ضریب تغییرات برابر با کدام گزینه است؟

الف. ۲ ب. ۲۰۰ ج. ۵ د. ۵۰

۳- ۱۲. در چه حالتی نمره‌ی استاندارد برابر با یک می‌شود؟

الف. واریانس آن برابر یک باشد.

ب. اختلاف از میانگین یک باشد.

ج. نمره‌ی اصلی برابر با مجموع انحراف معیار و میانگین باشد.

د. نمره‌ی اصلی برابر با دو برابر میانگین باشد.

۳- ۱۳. مجموع دو کران بازه‌ی چیشف، که حداقل ۷۵ درصد داده‌ها به آن تعلق دارد چقدر است؟

الف. صفر ب. $2\bar{x}$ ج. $4s$ د. $\frac{1}{4}s$

۳- ۱۴. اگر میانگین داده‌ها برابر ۸ باشد، براساس قضیه‌ی چیشف، کدام یک از بازه‌ها حداقل ۷۵ درصد داده‌ها را در بر دارد؟

الف. $(4, 12)$ ب. $(6, 12)$ ج. $(4, 10)$ د. بستگی به انحراف معیار دارد

۳- ۱۵. در کاربرد قضیه‌ی چیشف، برای تعیین حداقل درصدهایی که مجموعه‌ی داده‌ها در یک فاصله قرار گیرند،

الف. نیاز به محاسبه‌ی جدول فراوانی است.

ب. باید بزرگترین و کوچکترین مشاهدات را بیابیم.

ج. میانگین و انحراف معیار مشاهدات را بدانیم.

د. این محدودیت که k یک عدد طبیعی است باید رعایت شود.

۳- ۱۶. براساس تعریف نیم‌برد میان‌چارکی،

الف. با افزودن عددی به مشاهدات، به نیم‌بردمیان‌چارکی نیز افزوده می‌شود.

ب. تغییرات در ۷۵ درصد مشاهدات در محاسبه‌ی آن تأثیری نمی‌گذارد.

ج. میانگین انحراف چارک اول و چارک سوم از هر عددی است.

د. از روی میانه نیز می‌توان آن را محاسبه کرد.

۳-۱۷. $Q_1 = 8$ و $Q_3 = 13$ ، مقدار ضریب تغییر چارکی برابر کدام گزینه است؟

الف. ۴/۲ ب. ۴۲ ج. ۲۳/۸ د. ۱۰/۵

۳-۱۸. مطابق جدول آماری به شرح ذیل، مقدار چارک اول برابر با کدام گزینه است؟

رده‌ها	فراوانی	فراوانی تجمعی
۲۵-۴۵	۸	۸
۴۵-۶۵	۹	۱۷
۶۵-۸۵	۱۱	۲۸
۸۵-۱۰۵	۵	۳۳
۱۰۵-۱۲۵	۷	۴۰
۱۲۵-۱۴۵	۱۰	۵۰

الف. ۴۵ ب. ۶۵ ج. ۵۲/۲۲ د. ۳۲/۲۲

فصل چهارم

معیارهای ارزیابی منحنی فراوانی

هدف کلی

آموزش معیارهای ارزیابی منحنی فراوانی‌ها که شامل محاسبه‌ی معیار چولگی و کشیدگی، و شناسایی داده‌های دورافتاده

هدف‌های یادگیری

خواننده پس از یادگیری این فصل باید بتواند:

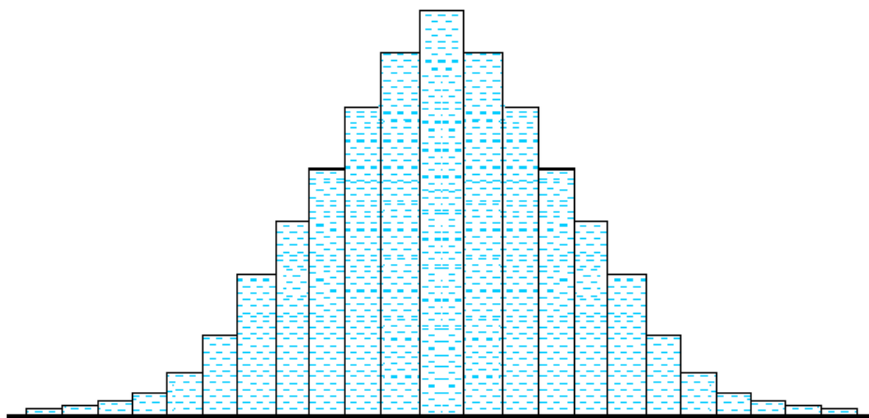
- بر اساس نمودارهای توزیع فراوانی‌ها، توصیفی از منحنی فراوانی‌ها ارائه دهد.
- مفهوم چولگی و انواع آن را شرح دهد.
- مفهوم کشیدگی را شرح دهد.
- انواع ضرایب چولگی را محاسبه کند و نوع چولگی توزیع فراوانی را تعیین کند.
- گشتاورهای مرکزی از هر مرتبه‌ای را محاسبه کند.
- ضریب کشیدگی یا برجستگی را به‌دست آورد.
- نمودار جعبه - خط را ترسیم کند.
- از روی نمودار جعبه - خط تقارن توزیع فراوانی را شرح دهد.
- از روی نمودار جعبه - خط مشاهدات دور افتاده را شناسایی کند.
- مقادیر معیارهای چولگی و کشیدگی برای منحنی نرمال را نشان دهد.

مقدمه

تاکنون انواع معیارهایی را که به نحوی «مرکز» داده‌ها و «تغییرپذیری» یا «پراکندگی» داده‌ها را توصیف می‌کردند، شرح دادیم. در واقع نمی‌توان حدی برای تعداد راه‌های توصیف داده‌ها قائل شد، زیرا محققان و پژوهشگران مرتباً روش‌های جدیدی برای توصیف مشخصه‌های عددی داده‌های مورد مطالعه‌ی خود در مسائل معین به وجود می‌آورند. در این فصل، به اختصار، به توصیف شکل کلی توزیع فراوانی داده‌ها می‌پردازیم. منظور از توزیع فراوانی داده‌ها، نحوه‌ی توزیع و پخش داده‌ها در قالب یک قانون است که می‌تواند یک نمودار یا یک تابع (منحنی) باشد.

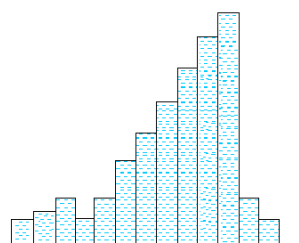
۴-۱ معیارهای ویژه‌ی نمودارهای داده‌ها

توزیع‌های فراوانی می‌توانند هر شکل و وضعی داشته باشند، اغلب توزیع‌هایی را که در عمل با آنها برخورد می‌کنیم، می‌توان با استفاده از یکی از چند نوع استاندارد توزیع‌های فراوانی به نحو نسبتاً رضایت‌بخشی توصیف کرد. مهم‌ترین توزیع فراوانی، توزیع زنگ شکل متقارن است که در نمودار ۴-۱ نشان داده شده است. در واقع دلایل نظری برای اینکه چرا بسیاری از توزیع‌ها از این الگو تبعیت می‌کنند، وجود دارد.

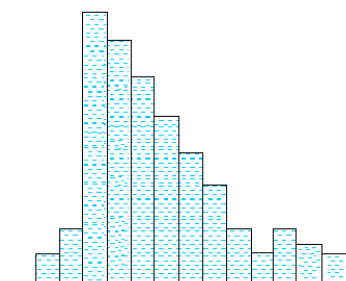


نمودار ۴-۱. توزیع زنگ شکل متقارن

دو توزیع دیگر در نمودارهای ۴-۲ و ۴-۳ را هنوز هم می‌توان با اندک مسامحه زنگ شکل نامید، اما این توزیع‌ها دیگر متقارن نیستند. توزیع‌های نظیر اینها که «دمچه» ای در یک طرف دارند، چوله نامیده می‌شوند؛ اگر دمچه در سمت چپ باشد، می‌گوییم که توزیع چولگی منفی دارد (نمودار ۴-۲) و اگر دمچه در راست باشد، می‌گوییم توزیع چولگی مثبت دارد (نمودار ۴-۳). توزیع‌های درآمدها یا دستمزدها اغلب به دلیل وجود تعدادی با مقادیر نسبتاً بزرگ، چولگی مثبت دارند. در این توزیع‌ها، مقادیر کوچک قادر به خنثی کردن اثر مقادیر بسیار بزرگ نیستند.



نمودار ۴-۲. توزیع با چولگی منفی



نمودار ۴-۳. توزیع با چولگی مثبت

راه‌های گوناگونی برای اندازه‌گیری میزان چولگی وجود دارد. یک روش نسبتاً آسان، بر این واقعیت مبتنی است که برای یک توزیع زنگ شکل کاملاً متقارن، از نوع توزیع توصیف شده در نمودار ۴-۱، مقادیر میانگین و میانه برهم منطبق‌اند. چون وجود تعدادی مقدار نسبتاً بزرگ که مقادیر کوچک متناظر، قادر به جبران آنها نیستند،

موجب می‌شود که میانگین بزرگتر از میانه باشد (و وجود مقادیر نسبتاً کوچک که مقادیر نسبتاً بزرگ نظیر قادر به جبران آنها نیستند موجب می‌شود که میانگین کوچکتر از میانه باشد). به‌طور معمول، در چولگی مثبت:

$$Md < Mo < \bar{x}$$

و در چولگی منفی:

$$\bar{x} < Mo < Md$$

برقرار است.

در واقع الگوی ایده‌آلی که نمودار ۴-۱ به صورت بافت‌نگار از آن پیروی می‌کند را می‌توان در قالب یک منحنی پیوسته با فرمول (۴-۱) به نام توزیع نرمال ارائه داد:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad x \in R, \quad (4-1)$$

که در آن پارامترهای μ و σ به ترتیب عدد حقیقی، و عددی مثبت هستند که با معلوم بودن آنها فرمول (۴-۱) کاملاً تعیین می‌شود. از این توزیع در فرمول (۳-۴) با عنوان دستور تجربی، برای تعیین درصد داده‌هایی که درون بازه‌هایی با طول $2ks$ قرار می‌گیرند، استفاده کرده‌ایم. تغییر در هیچ‌یک از دو کمیت μ و σ تأثیری در چوله‌شدن منحنی نرمال نمی‌گذارد، بلکه ممکن است منحنی را روی محور طول‌ها جابه‌جا کند یا ارتفاع قله‌ی منحنی را افزایش یا کاهش دهد. شرح ویژگی‌های این توزیع نرمال در درس «احتمال یک» ارائه می‌شود. در ادامه، فقط لازم به ذکر است که معیارهایی برای محاسبه‌ی عدم تقارن (چولگی) و همچنین برجستگی مد (کشیدگی) از مقایسه توزیع داده‌ها با توزیع نرمال حاصل می‌شود.

می‌توانیم از این روابط بین میانگین و میانه و نما استفاده کرده و معیارهای ساده‌ای برای میزان چولگی توزیع، تعریف کنیم. اولین معیار ضریب اول چولگی پیرسن (یا ضریب مد چولگی) نامیده می‌شود و مقدار آن چنین است:

$$SK_{P_1} = \frac{\bar{x} - Mo}{s}. \quad (4-2)$$

دومین معیار، ضریب دوم چولگی پیرسن (یا ضریب میانه چولگی) نامیده می‌شود، عبارت است از:

$$SK_{P_2} = \frac{3(\bar{x} - Md)}{s}. \quad (3-4)$$

برای یک منحنی کاملاً متقارن، مقدار SK ها برابر صفر است و در حالت کلی SK_P بین -3 و $+3$ تغییر می‌کند.

معیاری دیگری که از گشتاور مرکزی سوم به دست می‌آید، ضریب گشتاوری چولگی پیرسن نامیده می‌شود، برابر است با:

$$SK_P = \frac{m_3}{s^3}, \quad (4-4)$$

که در آن:

$$m_3 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3$$

گشتاور مرکزی مرتبه‌ی سوم، و s انحراف معیار نمونه است.

مثال ۴-۱. مطابق مثال ۲-۳، مربوط به میزان اکسید سولفور، مقدار چولگی را به دست آورید.

حل. دیدیم که

$$\bar{x} = 18.85, \quad Md = 19.03, \quad s = 5.55, \quad Mo = 19.27.$$

همچنین با قدری محاسبه‌ی گشتاور مرکزی مرتبه‌ی سوم، $m_3 = -16.48$ به دست می‌آوریم. بنابراین:

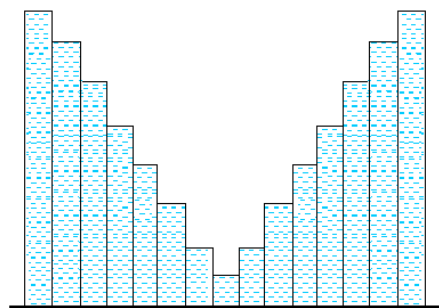
$$SK_{P_1} = \frac{(18.85 - 19.27)}{5.55} = -0.076$$

$$SK_{P_2} = \frac{3(18.85 - 19.03)}{5.55} = -0.01$$

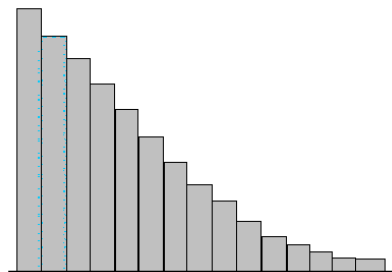
$$SK_P = \frac{-16.48}{(5.55)^3} = -0.094$$

ملاحظه می‌شود که مقدار سه معیار بسیار کوچک است و لذا می‌توان توزیع را تقریباً متقارن دانست.

دو نوع دیگر از توزیع‌هایی که در برخی مدل‌های کاربردی پیش می‌آیند، توزیع‌های U شکل و J معکوس شکل است که در نمودارهای ۴-۴ و ۴-۵ نشان داده شده‌اند. همان‌گونه که در نمودار این توزیع‌ها آشکار است، نام این توزیع‌ها تا حدی شکل توزیع‌ها را معین می‌کنند.



نمودار ۴-۴. توزیع U شکل



نمودار ۴-۵. توزیع J معکوس شکل

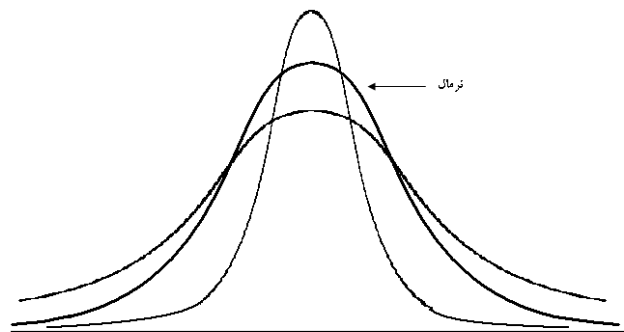
معیار دیگری در نمودارهای آماری مطرح است که به معیار کشیدگی یا برجستگی منحنی توجه دارد. معمولاً این معیار از مقایسه با منحنی نرمال به‌دست می‌آید. برجستگی به‌وسیله‌ی فرمول (۴-۵) محاسبه می‌شود:

$$K = \frac{m_4}{s^4} - 3 \quad (4-5)$$

که در آن

$$m_i = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^i$$

گشتاور مرکزی مرتبه‌ی چهارم و s انحراف معیار نمونه است. چون مقدار $\frac{m_i}{s^i}$ برای توزیع نرمال به عدد ۳ نزدیک است، معیار برجستگی را از تفاوت دو کمیت محاسبه می‌کنند. بسته به آنکه این کمیت مثبت و یا منفی باشد به ترتیب برجستگی بیشتر و یا کمتر از منحنی نرمال خواهند داشت.



شکل ۴-۶. وضعیت کشیدگی یا برجستگی منحنی‌ها نسبت به منحنی نرمال

مثال ۴-۲. مطابق مثال ۲-۳، مربوط به میزان اکسید سولفور، مقدار کشیدگی را به دست آورید.

حل. با محاسبه‌ای نسبتاً طولانی، مقدار گشتاور مرکزی چهارم، $m_4 = 1472,926$ به دست می‌آید. پس

$$K = \frac{1472,926}{(0,55)^4} - 3 = -1,65$$

این مقدار منفی حاکی از پایین‌تر قرار گرفتن قله‌ی منحنی نسبت به منحنی نرمال است. منحنی دارای پخی قابل ملاحظه‌ای است.

۴-۲ نمودار جعبه - خط و ارزیابی فراوانی‌ها

نمودار جعبه‌ای که در بخش ۳-۵ معرفی شد، نموداری است که به کمک معیارهای مرکزی، موقعیت مجموعه‌ی داده‌ها را به شکلی بسیار گویا و مفید ارائه می‌دهد. اینک با قدری تغییر در نمودار جعبه‌ای، نمودار جعبه - خط را ارائه می‌دهیم که از آن به عنوان ابزاری برای تفسیر میزان مرکزیت، پراکندگی، و چولگی داده‌ها استفاده خواهیم کرد. برای بررسی دقیق‌تر، مجموعه‌ی داده‌ها و همچنین توجه ویژه به نقاط پرت، «مرزهای داخلی بالا و پایین» و «مرزهای خارجی بالا و پایین» را تعریف می‌کنیم.

$$\begin{aligned} I_U &= Q_3 + 3 \times IQR = Q_3 + 1.5(Q_3 - Q_1) && \text{مرز داخلی بالا} \\ E_U &= Q_3 + 6 \times IQR = Q_3 + 3(Q_3 - Q_1) && \text{مرز خارجی بالا} \\ I_L &= Q_1 - 3 \times IQR = Q_1 - 1.5(Q_3 - Q_1) && \text{مرز داخلی پایین} \\ E_L &= Q_1 - 6 \times IQR = Q_1 - 3(Q_3 - Q_1) && \text{مرز خارجی پایین} \end{aligned}$$

اگر بزرگترین و کوچکترین داده، از مرزهای داخلی بالا و پایین کوچکتر باشند آنگاه خط توپری از این نقاط به میانه‌ی مستطیل وصل می‌شود، در غیر این صورت مرزهای داخلی بالا و پایین را به صورت خط منقطع به وسط مستطیل اتصال می‌دهیم. این خطوط که در برخی متون آن‌را «ویسکر» می‌نامند، از چارک‌ها شروع و به نقاط مرزهای داخلی ختم می‌شوند.

داده‌ای که بزرگتر (کوچکتر) از مرز داخلی بالا (پایین) و یا کوچکتر از مرز خارجی بالا (پایین) قرار گیرد را «داده‌ی پرت ضعیف» و داده‌ای که بزرگتر (کوچکتر) از مرز خارجی بالا (پایین) قرار گیرد را «داده‌ی پرت قوی» می‌نامند و به ترتیب آن‌را در شکل با نماد «○» و «●» نمایش می‌دهیم.

از روی نمودار جعبه - خط، نتیجه‌ی تفسیر دیداری را بدین شرح گزارش می‌کنیم:
تقریباً متقارن: اگر میانه نزدیک وسط مستطیل قرار گیرد، توزیع داده‌ها تقریباً متقارن است.

چولگی مثبت (منفی): در وضعیتی که میانه در قسمت بالای (پایین) مستطیل قرار گیرد، توزیع داده‌ها چولگی مثبت (منفی) دارد.

متقارن: اگر خطوط دو طرف مستطیل (ویسکرها) تقریباً برابر باشند، توزیع داده‌ها به حالت متقارن نزدیک است و در صورت نامساوی بودن خطوط، تأییدی دیگر بر چولگی توزیع داده‌ها است.

پراکندگی چندجامعه: اگر نمودار جعبه‌ای به‌طور همزمان برای چند جامعه ترسیم شود، جامعه‌ای که دارای طول مستطیل‌های بزرگتری باشد، پراکندگی بیشتری دارد.

داده‌های پرت قوی و ضعیف: با تعیین نقاط خارج از مرزهای تعریف‌شده، به سادگی می‌توان این‌گونه نقاط را مشخص کرد.

مثال ۴-۳. برای داده‌های زیر، نمودار جعبه - خط را رسم نموده و تفسیر توزیع فراوانی را انجام دهید.

۴/۳, ۶/۱, ۴/۰, ۴/۵, ۴/۴, ۴/۹, ۵/۸, ۴/۷, ۴/۱, ۴/۶, ۴/۴, ۴/۳, ۴/۸, ۴/۴, ۴/۲, ۴/۵, ۴/۴

حل. ابتدا هفده داده را مرتب و پنج مقدار لازم برای رسم نمودار جعبه - خط، نمودار ۴-۷ را تعیین می‌کنیم.

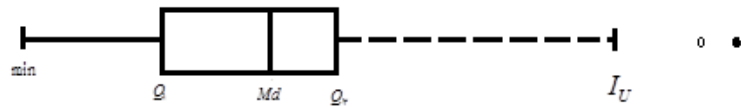
۴/۰, ۴/۱, ۴/۲, ۴/۳, ۴/۳, ۴/۴, ۴/۴, ۴/۴, ۴/۴, ۴/۵, ۴/۵, ۴/۶, ۴/۷, ۴/۸, ۴/۹, ۵/۸, ۶/۱
min Q_1 Md Q_3 max

مقادیر مرزهای داخلی و خارجی عبارت‌اند از:

$$\begin{aligned} I_U &= 4.7 + 1.5(4.7 - 4.3) = 5.3 && \text{مرز داخلی بالا} \\ E_U &= 4.7 + 3(4.7 - 4.3) = 5.9 && \text{مرز خارجی بالا} \\ I_L &= 4.3 - 1.5(4.7 - 4.3) = 3.7 && \text{مرز داخلی پایین} \\ E_L &= 4.3 - 3(4.7 - 4.3) = 2.1 && \text{مرز خارجی پایین} \end{aligned}$$

داده‌های پرت قوی و ضعیف، که به‌ترتیب در سمت راست و چپ، به‌شرح ذیل قرارگرفته‌اند. شکل توزیع فراوانی این داده‌ها، دمی بلند در سمت راست را نشان خواهد داد. بنابراین، به‌طور آشکار، چولگی راست در این توزیع وجود دارد. (تمرین ۴

$\frac{4}{0}, \frac{4}{1}, \frac{4}{2}, \frac{4}{3}, \frac{4}{3}, \frac{4}{4}, \frac{4}{4}, \frac{4}{4}, \frac{4}{4}, \frac{4}{5}, \frac{4}{5}, \frac{4}{6}, \frac{4}{7}, \frac{4}{8}, \frac{4}{9}, \frac{5}{8}, \frac{6}{8}, \frac{6}{1}$
 min Q_1 Med Q_3 max



نمودار ۷-۴. نمودار جعبه - خط برای داده‌های مثال ۴-۳

خلاصه‌ی فصل چهارم

یکی از ویژگی‌های منحنی توزیع فراوانی، متقارن بودن است که در اکثر پدیده‌های مورد مطالعه، حاصل نمی‌شود. چولگی یا عدم تقارن در دو حالت مثبت و منفی ممکن است پدیدار شود. برای محاسبه‌ی چولگی، براساس میانگین، میانه و مد سه نوع ضریب چولگی داریم:

$$SK_{P_1} = \frac{\bar{x} - Mo}{s}$$

$$SK_{P_2} = \frac{3(\bar{x} - Md)}{s}$$

$$SK_P = \frac{m_r}{s^3}$$

ویژگی دیگری که در منحنی توزیع فراوانی مورد ارزیابی قرار می‌گیرد، برجستگی است که با مقایسه‌ی منحنی توزیع فراوانی نرمال، معیار زیر مورد استفاده قرار می‌گیرد:

$$K = \frac{m_i}{s^4} - 3$$

استفاده از نمودار جعبه - خط برای تعیین تقارن و شناسایی داده‌های دورافتاده نیز به عنوان یک ابزار سودمند مورد بحث قرار گرفت.

تمرین

۴-۱. برای ۱۲۰ مشاهده، اطلاعات زیر به‌دست آمده است:

الف. میانگین، میانه، واریانس را به‌دست آورید.

ب. چارک‌های اول و سوم را محاسبه کنید.

ج. مقدار چولگی را به دست آورید.

د. محاسبات را مجدداً با نرم افزار به دست آورید.

L_U	f
۴۴/۵_۵۴/۵	۵
۵۴/۵_۶۴/۵	۴۵
۶۴/۵_۷۴/۵	۴۳
۷۴/۵_۸۴/۵	۱۹
۸۴/۵_۹۴/۵	۷
۹۴/۵_۱۰۴/۵	۱

۴-۲. طول عمر ۱۰۰ باتری خودرو به ترتیب دارای میانگین، میانه و انحراف استاندارد

۳/۵، ۳/۴۸، و ۱/۶۵ سال است. ضریب چولگی پیرسن را محاسبه کنید.

۴-۳. برای مشاهدات زیر، نمودار جعبه-خط را رسم کنید. توضیح دهید که توزیع

فراوانی از نظر چولگی، در چه وضعیتی است؟

۱ ۳ ۲ ۴ ۵ ۱ ۲ ۴ ۳ ۲ ۵ ۸ ۹ ۱ ۲

۴-۴. برای داده‌های مثال ۴-۳ ضرایب چولگی را به دست آورید.

خودآزمایی چهارگزینه‌ای

۴-۱. منفی شدن ضریب چولگی نوع دوم پیرسن، به علت است.

الف. $\bar{x} < Md$ ب. $\bar{x} < Mo$ ج. $\bar{x} > Md$ د. $\bar{x} > Mo$

۴-۲. هر قدر میزان معیار کشیدگی بزرگ باشد،

الف. پراکندگی کمتر است. ج. میانگین بزرگتر است.

ب. پراکندگی بیشتر است. د. میانگین کوچکتر است.

۴-۳. اگر $Mo = ۴$ ، $\bar{x} = ۵$ ، و $s = ۱۶$ ، آنگاه ضریب چولگی نوع دوم پیرسن برابر با

کدام گزینه است؟

الف. $\frac{۱}{۱۶}$ ب. $\frac{۱}{۴}$ ج. $\frac{۳}{۴}$ د. $\frac{۳}{۱۶}$

- ۴-۴. اگر میانه و میانگین برابر باشند، آنگاه.....
- الف. توزیع متقارن است.
- ب. در صورت صفر بودن این دو مقدار، توزیع متقارن است.
- ج. ضریب چولگی پیرسن مثبت می‌شود.
- د. تعیین شکل توزیع به مقدار مد بستگی پیدا می‌کند.
- ۴-۵. اگر گشتاور مرکزی مرتبه‌ی چهارم برابر ۲۵، و انحراف معیار برابر با ۴ باشد، آنگاه مقدار معیار کشیدگی کدام گزینه است؟
- الف. ۲/۹- ب. ۳- ج. ۰/۰۹۷- د. ۰/۰۹۷
- ۴-۶. در نمودار جعبه-خط از کدام گزینه استفاده نمی‌شود؟
- الف. میانگین ب. میانه ج. چارک اول د. چارک سوم

فصل پنجم

روش‌های اساسی شمارش

هدف کلی

آشنایی با روش‌هایی برای شمارش تعداد عناصر مجموعه‌ها و کاربرد آنها در تعیین ضرایب جملات در بسط دوجمله‌ای

هدف‌های یادگیری

- خواننده پس از یادگیری این فصل باید بتواند:
- دو اصل مهم از اصول شمارش یعنی «اصل جمع» و «اصل ضرب» را توضیح دهد.
- مفهوم جایگشت را شرح دهد.
- فاکتوریل را محاسبه کند و آن را برای محاسبه‌ی تعداد جایگشت‌ها به کار گیرد.
- مفهوم ترتیب را بیان نماید و فرمول محاسبه‌ی آن را اثبات کند.
- تعداد جایگشت دوری را محاسبه کند.
- تفاوت ترکیب با ترتیب را بیان کرده و فرمول محاسبه‌ی آن را اثبات کند.
- تعمیم و کاربرد فرمول محاسبه‌ی تعداد ترکیب را شرح دهد.
- به کمک اصول شمارش، ضرایب را در بسط دوجمله‌ای به دست آورد.
- محاسبه‌ی فرمول ترکیب به ازای مقادیر صحیح منفی را انجام دهد.
- برابری‌های مربوط به فرمول‌های ترکیب و ضرایب دوجمله‌ای را اثبات کند.

مقدمه

ساده‌ترین نوع محاسبه، شمارش تعداد اعضای یک مجموعه‌ی پایان‌دار است که انسان به طور طبیعی از طریق انگشتان دست یا اشیای دیگر، برای اولین بار به آن توجه کرده است. با این حال اگر این مجموعه‌ی پایان‌دار خیلی پرعضو باشد، به‌طوری که نتوانیم یک‌یک اعضای آن را نشان دهیم، عمل شمارش نه‌تنها از عهده‌ی انسان بلکه از عهده‌ی کامپیوتر هم ساخته نیست. در محاسبه‌ی احتمال‌ها که در فصل بعدی خواهیم دید، اغلب نیاز به شمارش حالات ممکن مختلف، بدون نیاز به نوشتن تمامی حالات، داریم. خوشبختانه به کمک اندیشه‌ی ریاضی می‌توان فرمول‌هایی ابداع کرد که از دشواری کار شمارش بکاهد. فن پیدا کردن این فرمول‌ها بر اصولی بدیهی به‌نام اصول شمارش بنا شده است.

۵-۱ اصول شمارش

در اینجا «عمل» نمایشگر هر نوع شیوه، فرایند، یا روش انتخاب است. فرض کنید عمل A را به n_1 طریق به‌نام‌های $x_1, x_2, \dots, x_{n_1}$ و عمل B را به n_2 راه با نام‌های $y_1, y_2, \dots, y_{n_2}$ بتوان انجام داد.

اصل اول شمارش. اگر انجام عمل L منوط به انجام کار A یا B باشد، آنگاه عمل L را می‌توان به $n_1 + n_2$ راه به‌نام‌های $x_1, x_2, \dots, x_{n_1}, y_1, y_2, \dots, y_{n_2}$ انجام داد. این اصل را «اصل جمع» می‌گویند و در آن تکیه بر روی «یا» است.

مثال ۵-۱. فرض کنید مسیر دانشکده به منزل (عمل L)، با تاکسی (عمل A) از سه راه یا با اتوبوس (عمل B) از دو راه امکان داشته باشد. بنابراین به پنج طریق با تاکسی یا با اتوبوس می‌توان از دانشکده به منزل رفت.

اصل دوم شمارش. اگر انجام عمل L طی دو مرحله متوالی انجام‌پذیر باشد، به‌طوری که در مرحله‌ی اول (عمل A) به n_1 راه و مرحله‌ی دوم (عمل B) به n_2 راه صورت گیرد، آنگاه عمل مزبور به $n_1 n_2$ راه انجام می‌گیرد.

برای توجیه این قضیه، زوج مرتب (x_i, y_i) را برآمدی تعریف می‌کنیم که وقتی مرحله‌ی اول به امکان x_i و مرحله دوم به امکان y_i منتج شود، رخ می‌دهد. در این صورت، مجموعه‌ی تمام برآمدهای ممکن، مرکب از $n_1 n_2$ زوج زیر است:

$$\begin{array}{cccc} (x_1, y_1), & (x_1, y_2), & \dots, & (x_1, y_{n_2}) \\ (x_2, y_1), & (x_2, y_2), & \dots, & (x_2, y_{n_2}) \\ \vdots & \vdots & \ddots & \vdots \\ (x_{n_1}, y_1), & (x_{n_1}, y_2), & \dots, & (x_{n_1}, y_{n_2}) \end{array}$$

این اصل را «اصل ضرب» می‌گویند و در آن تکیه بر روی «و» است.
مثال ۵-۲. فرض کنید از دانشکده به منزل (عمل L) با تاکسی از دانشکده تا پارک شهر (عمل A) از سه راه و با اتوبوس از پارک شهر تا منزل (عمل B) از دو راه امکان داشته باشد. بنابر این به شش طریق با تاکسی و با اتوبوس می‌توان از دانشکده به منزل رفت.

	x_1	x_2	x_3
y_1	(x_1, y_1)	(x_2, y_1)	(x_3, y_1)
y_2	(x_1, y_2)	(x_2, y_2)	(x_3, y_2)

تعمیم اصل اول شمارش. اگر در انجام عملی بتوانیم یکی از k عمل را انتخاب کنیم، به‌طوری که عمل اول بتواند به n_1 راه انجام شود، عمل دوم به n_2 راه انجام شود، و الی آخر، کل عمل به $n_1 + n_2 + \dots + n_k$ راه انجام می‌گیرد.

تعمیم اصل دوم شمارش. اگر عملی مرکب از k مرحله‌ی متوالی باشد، به‌طوری که مرحله‌ی اول بتواند به n_1 راه انجام شود، و برای هر یک از این راه‌ها، مرحله‌ی دوم بتواند به n_2 راه صورت گیرد، و برای هریک از راه‌های این دو مرحله‌ی نخستین، مرحله‌ی سوم بتواند به n_3 راه انجام گیرد و الی آخر، آنگاه کل عمل به $n_1 n_2 \dots n_k$ راه صورت می‌پذیرد.

مثال ۳-۵. برای دریافت اخبار رویدادهای محلی، پنج شبکه‌ی تلویزیونی، سه موج رادیویی، شش مجله‌ی الکترونیکی و دو روزنامه‌ی کاغذی وجود دارد. در کل می‌توان $۱۶ = ۲ + ۳ + ۵$ راه اخبار محلی را دریافت کرد.

مثال ۴-۵. از میان ۴ سوپ مختلف، ۳ نوع ساندویچ، ۵ نوع دسر و ۴ نوع نوشابه، چند ناهار مختلف که شامل یک سوپ، یک ساندویچ، یک دسر و یک نوشابه است می‌توان انتخاب کرد؟

حل. بنابر تعمیم اصل دوم شمارش جواب

$$۴ \times ۳ \times ۵ \times ۴ = ۲۴۰$$

است.

۵-۲ جایگشت

هر آرایش از اشیای متمایز را یک «جایگشت» می‌نامیم. فرض کنید n شیء متمایز داشته باشیم، می‌توانیم این اشیاء را به صورت‌های مختلف آرایش دهیم.

مثال ۵-۵. چند جایگشت از سه حرف a ، b ، و c وجود دارد؟

حل. آرایش‌های ممکن عبارت‌اند از:

$$abc \quad acb \quad bac \quad bca \quad cab \quad cba$$

بنابراین تعداد جایگشت‌های متمایز، ۶ تا است. می‌توانستیم بدون فهرست کردن جایگشت‌های مختلف، از «تعمیم اصل دوم شمارش» استفاده کنیم و بگوییم: در مکان اول برای جایگاه حرف سمت چپ (راست) سه انتخاب برای گزینش یک حرف وجود دارد، و سپس دو انتخاب برای مکان دوم موجود است. فقط یک حرف برای مکان سوم باقی می‌ماند، لذا تعداد کل جایگشت‌ها $۶ = ۳ \times ۲ \times ۱$ است.

با تعمیم استدلالی که در مثال ۵-۵ به کار رفت، درمی‌یابیم که n شیء متمایز را می‌توان به $۱ \times ۲ \times ۳ \times \dots \times (n-۲) \times (n-۱) \times n$ راه آرایش داد. این حاصل ضرب را با نماد $n!$ که « n فاکتوریل» خوانده می‌شود، نمایش می‌دهیم. پس:

$$۱! = ۱$$

$$۲! = ۲ \times ۱$$

$$3! = 3 \times 2 \times 1$$

⋮

$$n! = n \times (n-1) \times \cdots \times 2 \times 1$$

بنابر تعریف $0! = 1$.

قضیه ۱-۵. تعداد جایگشت‌های متمایز n شیء متمایز برابر $n!$ است.

مثال ۵-۶. تعداد حالت‌هایی که پنج نفر در یک صف قرار می‌گیرند، برابر با $5! = 60$ است.

۵-۳ ترتیب

هرگاه از n شیء متمایز، r تا را برگزیده، (با شرط $1 \leq r \leq n$)، و یک جایگشت تشکیل دهیم، حاصل، یک ترتیب r تایی از n شیء متمایز است. تعداد جایگشت‌های r تایی از n شیء متمایز را بدین طریق پیدا می‌کنیم: برای ساختن یک جایگشت r تایی از n شیء متمایز، باید r تا از این n شیء را یک‌به‌یک انتخاب کرده، مثلاً از چپ به راست، پهلوی هم قرار دهیم. این کار معادل با انجام پیایی r عمل $A_1, A_2, \dots, A_{r-1}, A_r$ به ترتیب، $n, n-1, \dots, n-r+2, n-r+1$ طریق است. بنابر تعمیم اصل دوم شمارش، تعداد جایگشت‌های r تایی از n شیء متمایز می‌شود:

$$n \times (n-1) \times \cdots \times (n-r+2) \times (n-r+1).$$

این حاصل ضرب را با ${}_nP_r$ نشان می‌دهند. بنا به تعریف ترتیب، تعداد جایگشت‌های صفرتایی از n شیء متمایز را برابر یک قرار می‌دهیم، یعنی ${}_nP_0 = 1$.

قضیه ۵-۲. تعداد ترتیب‌های r تایی از n شیء متمایز برای $r = 0, 1, 2, \dots, n$ عبارت‌است از

$${}_nP_r = \frac{n!}{(n-r)!}.$$

برهان.

برای $r = 1, 2, \dots, n$ داریم:

$${}_nP_r = n(n-1)(n-2)\cdots(n-r+1) \\ = \frac{n(n-1)(n-2)\cdots(n-r+1)(n-r)!}{(n-r)!} = \frac{n!}{(n-r)!}.$$

فرمول ${}_nP_r = n(n-1)\cdots(n-r+1)$ را نمی‌توان برای $r=0$ به کار برد، اما فرمول زیر را می‌توان به کار برد:

$${}_nP_0 = \frac{n!}{(n-0)!} = 1.$$

مثال ۵-۷. از فهرست نام ۲۴ عضو یک باشگاه، برای انتخاب رئیس، نایب‌رئیس، خزانه‌دار، و منشی ۴ نام استخراج می‌شود. به چند راه مختلف این کار را می‌توان انجام داد؟

حل. تعداد جایگشت‌های ۲۴ شیء متمایز که ۴ به ۴ اختیار می‌شوند برابر است با:

$${}_{24}P_4 = \frac{24!}{20!} = 24 \times 23 \times 22 \times 21 = 255,024.$$

به کمک نرم‌افزار «آر»، محاسبه‌ی شمارش ویژگی‌های مشخص امکان‌پذیر است. برای مثال در یک سری داده، تعداد مشاهداتی که کوچکتر، بزرگتر، و یا بین دو عدد قرار دارند قابل تعیین است. همچنین نسبت‌دادن یک آرایش تصادفی به اعداد، از توانمندی‌های این نرم‌افزار به شمار می‌آید. توضیحات بیشتر در این خصوص در فصل هفتم آورده شده است.

۵-۴ جایگشت دوری

جایگشت‌های اشیاء که روی دایره‌ای مرتب شده‌اند، جایگشت‌های دوری خوانده می‌شوند. اگر اشیاء متناظر در دو آرایش، اشیای قبل و بعد یکسانی داشته باشند، دو آرایش را مختلف تلقی نمی‌کنیم (و آنها را یک بار به حساب می‌آوریم). برای مثال اگر چهار نفر دور میزی گرد نشسته باشند، وقتی هر کس به صندلی سمت چپ خود منتقل شود، جایگشتی مختلف به وجود نمی‌آید. بنابراین اگر شخصی را به دلخواه در مکان ثابتی در نظر بگیریم، سه نفر دیگر را طی سه مرحله می‌توانیم بنشانیم که حاصل تعداد این سه مرحله، شش جایگشت دوری است.

با تعمیم استدلالی که در این مثال ارائه شد، نتیجه‌ای را که در قضیه‌ی ۵-۳ بیان شده است به دست می‌آوریم.

قضیه‌ی ۵-۳. تعداد جایگشت‌های n شیء متمایز که روی یک دایره مرتب شده‌اند برابر است با $(n-1)!$.

مثال ۵-۸. به چند طریق می‌توان پنج آگهی تبلیغاتی را دور یک میدان نصب کرد؟
حل. براساس تعداد جایگشت‌های دوری، پاسخ $۲۴ = ۴! = (۵-۱)!$ است.

۵-۵ ترکیب

اگر n شیء متمایز داشته باشیم و بخواهیم از این n شیء فقط r شیء را انتخاب کنیم به‌طوری که ترتیب r شیء مهم نباشد، مسلماً به صورت‌های مختلف می‌توانیم این r شیء را برگزینیم. هر یک از این صورت‌ها را یک ترکیب r تایی از n شیء می‌نامیم.

تعداد ترکیب‌های r تایی از r شیء، تنها یک ترکیب r تایی است. بنابراین تعداد ترکیب‌های r تایی از n شیء متمایز که معمولاً آن را با $\binom{n}{r}$ یا ${}_nC_r$ نشان می‌دهند، برابر می‌شود با:

$${}_nC_r = \binom{n}{r} = \frac{{}_nP_r}{r!} = \frac{n!}{r!(n-r)!}.$$

قضیه‌ی ۵-۴. تعداد ترکیب‌های r شیء از n شیء متمایز برای $r = 0, 1, 2, \dots, n$ عبارت‌است از:

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}.$$

مثال ۵-۹. در ۶ پرتاب یک سکه به چند طریق ۲ شیر و ۴ خط ظاهر می‌شوند؟
حل. می‌توان گفت در چند راه مختلف ۲ پرتاب که در آن شیر ظاهر می‌شود، وجود دارد:

$$\binom{6}{2} = \frac{6!}{2!4!} = 15$$

این نتیجه را می‌توانستیم با فرایندی پرزحمت‌تر با شمارش حالت‌های مختلف، یعنی

$$HHTTTT, HTHTTT, \dots$$

که در آن H نمایش شیر و T نمایش خط است به‌دست آوریم.

مثال ۵-۱۰. می‌خواهیم از یک گروه ۲۰ نفری یک شورای ۳ نفره را تشکیل دهیم. به چند طریق مختلف این کار امکان‌پذیر است؟

حل. تعداد حالت ممکن برابر است با:

$$\binom{20}{3} = \frac{20 \times 19 \times 18}{3 \times 2 \times 1} = 1140.$$

مثال ۵-۱۱. به چند طریق n گوی در n جعبه طوری توزیع می‌شوند که فقط جعبه‌ی شماره یک خالی باشد؟

حل. فقط جعبه‌ی شماره یک خالی باشد؛ یعنی n گوی در $n-1$ جعبه به‌طریقی توزیع شوند که هیچ جعبه‌ای خالی نباشند. بنابراین دقیقاً یکی از $n-1$ جعبه باید دارای دو گوی باشد و $n-2$ جعبه‌ی دیگر باید دقیقاً دارای یک گوی باشند. فرض کنید جعبه‌ی j ام، $j = 2, \dots, n$ دارای دو گوی است. پس برای انتخاب دو گوی داخل جعبه‌ی j ام، $\binom{n}{2}$ راه و $n-2$ گوی مانده به $(n-2)!$ راه می‌توانند در $n-2$ جعبه‌ی دیگر قرار گیرند. در نتیجه با $n-1$ جعبه، تعداد راه‌ها برابر $(n-1) \times \binom{n}{2} \times (n-2)!$ می‌شود.

در اغلب منابع، قواعد شمارش با مدل‌های «جعبه و مهره» بیان می‌شوند. (بهبودیان، ۱۳۹۰) فرض کنید تعداد n جعبه‌ی متمایز را بخواهیم با تعداد r مهره پر کنیم. بر حسب اینکه مهره‌ها متمایز یا غیرمتمایز باشند و در هر جعبه گذاشتن مهره‌ی مکرر مجاز یا غیر مجاز باشد، چهار مدل (جدول ۵-۱) داریم.

جدول ۱-۵. قواعد شمارش در چهار مدل جعبه و مهره

تعداد حالات	شرح حالت
n^r	۱- مهره‌ها متمایز و مهره‌ی مکرر مجاز
$\binom{n}{r} \times r! = \frac{n!}{(n-r)!}$	۲- مهره‌ها متمایز و مهره‌ی مکرر غیرمجاز $r \leq n$
$\binom{n}{r} \times 1 = \frac{n!}{r!(n-r)!}$	۳- مهره‌ها غیرمتمایز و مهره‌ی مکرر غیرمجاز $r \leq n$
$\binom{n+r-1}{r} = \frac{(n+r-1)!}{r!(n-1)!}$	۴- مهره‌ها غیرمتمایز و مهره‌ی مکرر مجاز

مثال ۱۲-۵. معادله‌ی

$$x + y + z = 5, \quad x, y, z \in \{0, 1, 2, 3, 4, 5\}$$

چند جواب دارد؟

حل. مانند آن است که بخواهیم پنج مهره‌ی غیرمتمایز را در سه ظرف توزیع کنیم به طوری که هیچ محدودیتی بر ظرفیت ظرف‌ها نداشته باشیم. در زیر وضعیت ظرف‌هایی که جایگاه سه مجهول و خط‌چین‌هایی که مرز بین ظرف‌ها را نشان می‌دهد، ارائه شده است:

$$\begin{array}{lcl}
 | \bullet \bullet \bullet \bullet \bullet | & | & | \quad x = 5, y = z = 0 \\
 | \bullet \bullet \bullet \bullet & | \bullet & | \quad x = 4, y = 1, z = 0 \\
 \vdots & & \\
 | & | & | \bullet \bullet \bullet \bullet \bullet | \quad x = y = 0, z = 5
 \end{array}$$

در واقع این جایگشت‌ها، آرایش‌های دو خط و پنج مهره‌ی غیرمتمایز است. بنابراین تعداد جواب این معادله با قراردادن در فرمول بند ۴ از جدول ۱-۵ به ازای $n = 3$ و $r = 5$ برابر است با:

$$\binom{3+5-1}{5} = \binom{7}{5} = \frac{7!}{5! \times 2!} = 7 \times 3 = 21.$$

۵-۶ تعمیم ترکیب

یک ترکیب r تایی از n شیء دارای این تعبیر است که مجموعه‌ای از n شیء را به دو زیرمجموعه‌ی r شیء و $n-r$ شیء افراز می‌کنیم. پس تعداد ترکیب‌های r تایی از n شیء یعنی تعداد راه‌هایی که می‌توان یک مجموعه‌ی n شیء را به دو زیرمجموعه‌ی r شیء و $n-r$ شیء افراز کرد. حال این تعبیر را تعمیم می‌دهیم. می‌خواهیم تعیین کنیم که یک مجموعه‌ی n شیء را به چند راه می‌توانیم به k زیرمجموعه که در اولی n_1 شیء، در دومی n_2 شیء، ... و در k امی n_k شیء وجود دارد، افراز کنیم.

قضیه ۵-۵. تعداد ترکیب‌های n شیء که n_1 تایی آنها از نوع اول، n_2 تایی آنها از نوع دوم، ...، n_k تایی آنها از نوع k ام است و $n_1 + n_2 + \dots + n_k = n$ ، برابر است با:

$$\binom{n}{n_1, n_2, \dots, n_k} = \frac{n!}{n_1! n_2! \dots n_k!}.$$

مثال ۵-۱۳. به چند طریق می‌توان ۲۰ دانشجو را در کلاسی که در ردیف اول آن ۵ صندلی، در ردیف دوم آن ۶ صندلی، و در ردیف سوم آن ۹ صندلی دارد، نشانند؟
حل. در واقع مجموعه‌ای ۲۰ نفری را به چند راه می‌توان به زیرمجموعه‌های ۵ نفری، ۶ نفری و ۹ نفری افراز کرد:

$$\binom{20}{5, 6, 9} = \frac{20!}{5! 6! 9!} = 189878780.$$

مثال ۵-۱۴. یک مرکز پلیس در شهری کوچک دارای ۱۰ افسر پلیس است. اگر برنامه‌ی مرکز این باشد که ۵ افسر وظیفه نظارت بر خیابان‌ها، ۲ افسر پاسداری از مرکز، و ۳ افسر به‌صورت ذخیره باشند، به چند حالت ممکن می‌توان ۱۰ افسر را به ۳ گروه تقسیم نمود؟

حل. تعداد حالت‌های ممکن برابر است با:

$$\frac{10!}{5! 2! 3!} = 2520.$$

۵-۷ ضرایب دوجمله‌ای

اگر n عدد صحیح مثبتی باشد و $(x+y)^n$ را جمله‌به‌جمله در هم ضرب کنیم، هر جمله حاصل ضربی از x ها و y ها است، به‌طوری که از هریک از n عامل $(x+y)$ ، یک x یا یک y در این جمله می‌آید. مثلاً بسط

$$\begin{aligned}(x+y)^3 &= (x+y)(x+y)(x+y) \\ &= xxx + xxy + xyx + xyy + yxx + yxy + yyx + yyy \\ &= x^3 + 3x^2y + 3xy^2 + y^3\end{aligned}$$

جملاتی به شکل x^3 , x^2y , xy^2 , y^3 را نتیجه می‌دهد. ضرایب این جملات ۱، ۳، ۳، ۱ هستند، و برای مثال ضریب xy^2 ، برابر $\binom{3}{2} = 3$ ، یعنی تعداد راه‌هایی است که می‌توان دو عاملی که y ها را می‌دهند انتخاب کرد. همین‌طور ضریب x^2y برابر $\binom{3}{1} = 3$ ، یعنی تعداد راه‌هایی است که می‌توان یک عامل فراهم‌کننده‌ی y را

انتخاب کرد، و ضرایب x^3 و y^3 به ترتیب $\binom{3}{0} = 1$ و $\binom{3}{3} = 1$ هستند.

کلی‌تر آن‌که اگر n عدد صحیح و مثبتی باشد و $(x+y)^n$ را جمله‌به‌جمله در هم ضرب کنیم، ضریب $x^{n-r}y^r$ برابر $\binom{n}{r}$ ، یعنی تعداد راه‌هایی است که می‌توان r عامل فراهم‌کننده‌ی y ها را انتخاب کرد. از این‌رو $\binom{n}{r}$ را ضریب دوجمله‌ای می‌نامیم. حال می‌توانیم قضیه‌ی زیر را بیان کنیم.

قضیه‌ی ۵-۶. برای هر عدد صحیح مثبت n ،

$$(x+y)^n = \sum_{r=0}^n \binom{n}{r} x^{n-r} y^r.$$

محاسبه‌ی ضرایب دوجمله‌ای را اغلب با استفاده از سه قضیه‌ی زیر می‌توان ساده کرد:

قضیه ۵-۷. برای هر عدد صحیح مثبت n ,

$$\binom{n}{r} = \binom{n}{n-r}, \quad r = 0, 1, 2, \dots, n.$$

برهان. می‌توانیم استدلال کنیم که وقتی زیرمجموعه‌ای از r شیء را از مجموعه‌ی n شیء متمایز انتخاب می‌کنیم، زیرمجموعه‌ای با $n-r$ شیء باقی می‌ماند، و بنابراین راه‌های باقی‌ماندن (یا انتخاب کردن) $n-r$ شیء برابر تعداد راه‌های انتخاب r شیء است. برای این که قضیه را به صورت جبری ثابت کنیم، می‌نویسیم:

$$\begin{aligned} \binom{n}{n-r} &= \frac{n!}{(n-r)! [n-(n-r)]!} \\ &= \frac{n!}{(n-r)! r!} = \frac{n!}{r! (n-r)!} = \binom{n}{r}. \end{aligned}$$

از قضیه ۵-۷ نتیجه می‌شود که اگر ضرایب دوجمله‌ای را وقتی n زوج است برای $r = 0, 1, \dots, \frac{n}{2}$ و وقتی n فرد است برای $r = 0, 1, \dots, \frac{n-1}{2}$ محاسبه کنیم، ضرایب دوجمله‌ای باقی‌مانده را می‌توان با استفاده از این قضیه به دست آورد.

مثال ۵-۱۵. به فرض اینکه $\binom{4}{0} = 1$, $\binom{4}{1} = 4$, $\binom{4}{2} = 6$, مقادیر $\binom{4}{3}$ و $\binom{4}{4}$ را بیابید.

حل. $\binom{4}{3} = \binom{4}{4-3} = \binom{4}{1} = 4$ و $\binom{4}{4} = \binom{4}{4-4} = \binom{4}{0} = 1$.

قضیه ۵-۸. برای هر عدد صحیح مثبت n ,

$$\binom{n}{r} = \binom{n-1}{r} + \binom{n-1}{r-1}; \quad r = 1, 2, \dots, n-1.$$

برهان. در $(x+y)^n$ قرار می‌دهیم $x=1$ ، و می‌نویسیم:

$$\begin{aligned} (1+y)^n &= (1+y)(1+y)^{n-1} \\ &= (1+y)^{n-1} + y(1+y)^{n-1}. \end{aligned}$$

ضریب y^r در $(1+y)^n$ را با ضریب y^r در $(1+y)^{n-1} + y(1+y)^{n-1}$ برابر می‌گیریم. چون ضریب y^r در $(1+y)^n$ برابر با $\binom{n}{r}$ و ضریب y^r در $(1+y)^{n-1} + y(1+y)^{n-1}$ برابر مجموع ضریب y^r در $(1+y)^{n-1}$ یعنی $\binom{n-1}{r}$ و ضریب y^{r-1} در $(1+y)^{n-1}$ یعنی $\binom{n-1}{r-1}$ است، به دست می‌آوریم.

$$\binom{n}{r} = \binom{n-1}{r} + \binom{n-1}{r-1}$$

که برهان را کامل می‌کند.

به روشی دیگر، هر یک از n شیء را اختیار می‌کنیم، اگر یک شیء مشخص را در میان r شیء منظور نکنیم $\binom{n-1}{r}$ راه برای انتخاب r شیء وجود دارد، اگر آن شیء مشخص را در میان r شیء منظور کنیم $\binom{n-1}{r-1}$ راه برای انتخاب $r-1$ شیء دیگر وجود دارد. بنابراین $\binom{n-1}{r} + \binom{n-1}{r-1}$ راه برای انتخاب r شیء موجود است. برای بیان سومین قضیه درباره‌ی ضرایب دوجمله‌ای، تعریف می‌کنیم که هرگاه n ، یک عدد صحیح مثبت و r ، عددی صحیح و مثبت و بزرگتر از n باشد، $\binom{n}{r} = 0$. (بدیهی است امکان ندارد بتوانیم زیرمجموعه‌ای از مجموعه‌ای انتخاب کنیم که تعداد عناصرش از خود مجموعه بیشتر باشد).

قضیه ۹-۵. برای هر k, m و n صحیح مثبت،

$$\sum_{r=0}^k \binom{m}{r} \binom{n}{k-r} = \binom{m+n}{k}.$$

برهان. این قضیه را با برابر گرفتن ضرایب y^k در عبارت‌های دو طرف معادله‌ی

$$(1+y)^{m+n} = (1+y)^m (1+y)^n$$

ثابت می‌کنیم. ضریب y^k در $(1+y)^{m+n}$ برابر با $\binom{m+n}{k}$ است و ضریب y^k در $(1+y)^m(1+y)^n = \left[\binom{m}{0} + \binom{m}{1}y + \dots + \binom{m}{m}y^m \right] \times \left[\binom{n}{0} + \binom{n}{1}y + \dots + \binom{n}{n}y^n \right]$ مجموع حاصل ضرب‌هایی است که از ضرب جمله‌ی ثابت اولین عامل در ضریب y^k عامل دوم، ضرب ضریب y اولین عامل در ضریب y^{k-1} عامل دوم، ... و ضرب ضریب y^k اولین عامل در جمله‌ی ثابت عامل دوم به دست می‌آید. بنابراین ضریب y^k در $(1+y)^m(1+y)^n$ برابر است با:

$$\binom{m}{0}\binom{n}{k} + \binom{m}{1}\binom{n}{k-1} + \dots + \binom{m}{k}\binom{n}{0} = \sum_{r=0}^k \binom{m}{r}\binom{n}{k-r}$$

و برهان کامل می‌شود.

با استفاده از فرمول تعمیم ترکیب، می‌توان ضرایب بسط $(x_1 + x_2 + \dots + x_n)^n$ را تعیین کرد. ضریب چندجمله‌ای، جمله‌ی $x_1^{r_1} x_2^{r_2} \dots x_k^{r_k}$ در بسط $(x_1 + x_2 + \dots + x_n)^n$ برابر است با:

$$\binom{n}{r_1, r_2, \dots, r_k} = \frac{n!}{r_1! r_2! \dots r_k!},$$

در صورتی که شرط $n = r_1 + r_2 + \dots + r_k$ برقرار باشد.

مثال ۵-۱۶. در بسط $(x_1 + x_2 + x_3)^6$ ضریب $x_1^3 x_2^2 x_3^1$ چیست؟

حل. چون $n=6$ ، $r_1=3$ ، $r_2=2$ ، $r_3=1$ ، پس

$$\frac{6!}{3! 2! 1!} = 60.$$

خلاصه‌ی فصل پنجم

در این فصل ابتدا دو اصل مهم شمارش یعنی «اصل جمع» و «اصل ضرب» بیان شد. با توجه به مفهوم جایگشت، نشان داده شد که تعداد جایگشت‌های n شیء متمایز، برابر با $n!$ است. تعداد ترتیب‌ها و ترکیب‌های r شیء از n به ترتیب عبارت‌اند از:

ترکیب‌ها که در آن عناصر جامعه از k نوع تشکیل شده است، مطرح شد. با توجه به بسط دوجمله‌ای، و فرمول ضرایب آن، برابری‌های زیر اثبات شد:

$$(x+y)^n = \sum_{r=0}^n \binom{n}{r} x^{n-r} y^r, \quad n=1,2,\dots \quad ۱-$$

$$\binom{n}{r} = \binom{n}{n-r}, \quad r=0,1,2,\dots,n. \quad ۲-$$

$$\binom{n}{r} = \binom{n-1}{r} + \binom{n-1}{r-1}, \quad r=1,2,\dots,n-1. \quad ۳-$$

$$\sum_{r=0}^k \binom{m}{r} \binom{n}{k-r} = \binom{m+n}{k}, \quad m,n,k \in N \quad ۴-$$

تمرین

۱- ۵. اگر A یک مجموعه با عناصر $x_1, \dots, x_{n_1}$ و B مجموعه‌ای دیگر با عناصر $y_1, \dots, y_{n_2}$ باشند، و عمل L با یک عضو از A یا از B انجام گیرد آنگاه عمل L همتای کدام عمل روی دو مجموعه‌ی A و B است؟

۲- ۵. چند عدد یک‌رقمی که مضرب ۲ یا ۷ هستند، وجود دارد؟

۳- ۵. اگر A یک مجموعه با عناصر $x_1, \dots, x_{n_1}$ و B مجموعه‌ای دیگر با عناصر $y_1, \dots, y_{n_2}$ باشد، و عمل L منوط به برقراری یک عضو از A و یک عضو از B تعریف شود، آنگاه عمل L همتای کدام عمل روی دو مجموعه‌ی A و B است؟

۴- ۵. به چند طریق می‌توانید یکی از سه جفت کفش را با یکی از چهار جفت جوراب بپوشید؟

۵- ۵. اعضای شورای دانشجویی یک دانشگاه متشکل از ۳ نفر دانشجوی سال اول، ۴ نفر دانشجوی سال دوم، ۵ نفر دانشجوی سال سوم و ۲ نفر دانشجوی سال چهارم هستند. اگر بخواهیم یک شورای فرعی ۴ نفری را که دانشجویان سال‌های مختلف در آن هستند از این شورا انتخاب کنیم، چند شورای فرعی می‌توان انتخاب کرد؟

۶- ۵. چند پلاک نمره‌ی اتومبیل با هفت علامت که ۳ علامت اول آن از حروف الفبای لاتین و ۴ علامت آخر آن از اعداد صفر تا نه تشکیل شده است، می‌توان تهیه کرد؟

- ۵-۷. چند تابع را می‌توان روی n نقطه تعریف کرد، اگر هر تابع بتواند فقط مقادیر صفر و یک را اختیار کند؟
- ۵-۸. با ارقام ۱, ۲, ..., ۹ چند عدد نه‌رقمی که در آن ارقام تکراری نباشند، می‌توانید بسازید؟
- ۵-۹. یک کلاس نظریه‌ی احتمال دارای ۶ دانشجوی پسر و ۴ دانشجوی دختر است. پس از برگزاری یک امتحان، نمرات آنها مرتب می‌شوند. با فرض اینکه هیچ دو دانشجویی نمره‌ی یکسان کسب نکنند،
- الف. چند حالت ممکن برای مرتب کردن نمرات وجود دارد؟
- ب. اگر مرتب کردن دانشجویان پسر و دانشجویان دختر بین خودشان به‌طور جداگانه انجام گیرد، در این صورت چند حالت ممکن وجود دارد؟
- ۵-۱۰. شخصی ۱۰ کتاب دارد و می‌خواهد در قفسه‌ی کتابخانه خود قرار دهد. از ۱۰ کتاب، تعداد ۴ کتاب ریاضی، ۳ کتاب شیمی، ۲ کتاب تاریخ و یک کتاب زبان است. اگر این شخص بخواهد کتاب‌هایی با موضوع یکسان کنار هم قرار گیرند، آنگاه چند ترتیب ممکن برای قراردادن کتاب‌ها وجود دارد؟
- ۵-۱۱. در یک مسابقه‌ی شطرنج از ۱۰ بازیکن، ۴ نفر شیرازی، ۳ نفر تهرانی، ۲ نفر اصفهانی، و یک نفر مشهدی هستند. اگر نتیجه‌ی مسابقه براساس مرتب‌کردن شهر بازیکن‌ها اعلام شود، چند نتیجه ممکن وجود دارد؟
- ۵-۱۲. چند علامت مختلف را که هر کدام شامل قرارگرفتن ۹ پرچم در یک خط است می‌توان به وسیله‌ی ۴ پرچم سفید، ۳ پرچم قرمز، ۲ پرچم آبی تهیه کرد؟ پرچم‌های هم‌رنگ یکسان هستند.
- ۵-۱۳. یک شعبه‌ی محلی انجمن ریاضی به چند طریق می‌تواند برای سه گردهمایی مختلف، سه سخنران را تعیین کند، به شرطی که هر پنج سخنران برای هر یک از تاریخ‌های ممکن، حاضر به انجام سخنرانی باشند؟
- ۵-۱۴. چند جایگشت دوری از ۴ نفر که دور یک میز گرد نشسته‌اند وجود دارد؟
- ۵-۱۵. اگر در یک گروه شش‌تایی دو شیء کاملاً مشابه یکدیگر بودند، چند جایگشت دوری خواهیم داشت؟

۵-۱۶. از یک گروه متشکل از ۵ زن و ۷ مرد، چند شورای مختلف ۵ عضوی متشکل از ۲ زن و ۳ مرد می‌توان انتخاب کرد؟ اگر ۲ نفر از مردها با یکدیگر خصومت داشته باشند و نخواهند با هم در شورا انتخاب شوند، آنگاه چند شورا را می‌توان انتخاب نمود؟

۵-۱۷. در یک آزمون که شامل ۱۰ سؤال دوگزینه‌ای است به چند طریق می‌توان، با انتخاب تصادفی ۴ پاسخ صحیح داد؟

۵-۱۸. مجموعه‌ای از چهار شیء را به چند راه می‌توان به ۳ زیرمجموعه که به ترتیب شامل دو، یک، و یک شیء باشد افراز کرد؟

۵-۱۹. در فرمول قضیه‌ی ۵-۶، با قراردادن مقادیر مناسب به جای x و y نشان دهید

که: الف. $\sum_{r=0}^n \binom{n}{r} = 2^n$ ب. $\sum_{r=0}^n (-1)^r \binom{n}{r} = 0$ ج.

$$\sum_{r=0}^n \binom{n}{r} (a-1)^r = a^n$$

۵-۲۰. با به‌کاربردن مکرر قضیه‌ی ۵-۸ نشان دهید که:

$$\binom{n}{r} = \sum_{i=1}^{r+1} \binom{n-i}{r-i+a}$$

۵-۲۱. با به کار بردن قضیه‌ی ۵-۹ نشان دهید که:

$$\sum_{r=0}^n \binom{n}{r}^2 = \binom{2n}{n}$$

۵-۲۲. با قراردادن $x=1$ در رابطه‌ی قضیه‌ی ۵-۶، و سپس مشتق‌گیری از عبارت‌های دو طرف نسبت به y و بالاخره قراردادن $y=1$ ، نشان دهید که

$$\sum_{r=0}^n r \binom{n}{r} = n 2^{n-1}$$

۵-۲۳. اگر α ، عدد صحیح مثبت یا صفر نباشند، بسط دوجمله‌ای $(1+y)^\alpha$ به‌ازای

$-1 < y < 1$ به‌صورت سری نامتناهی

$$1 + \binom{\alpha}{1}y + \binom{\alpha}{2}y^2 + \binom{\alpha}{3}y^3 + \dots + \binom{\alpha}{r}y^r + \dots$$

تعریف می‌شود،

$$\binom{\alpha}{r} = \frac{\alpha(\alpha-1)\cdots(\alpha-r+1)}{r!}$$

این تعریف تعمیم یافته ضرایب دوجمله‌ای را برای محاسبه‌ی موارد زیر به کار برید:

الف. $\binom{\frac{1}{2}}{4}$ و $\binom{-3}{3}$.

ب. $\sqrt{5}$ ، با نوشتن $\sqrt{5} = 2\left(1 + \frac{1}{4}\right)^{\frac{1}{2}}$ و با استفاده از چهار جمله‌ی اول بسط دوجمله‌ای $\left(1 + \frac{1}{4}\right)^{\frac{1}{2}}$.

۵-۲۴. با رجوع به تعریف تعمیم یافته‌ی ضرایب دوجمله‌ای تمرین قبل، نشان دهید که

الف. $\binom{-1}{r} = (-1)^r$.

ب. به ازای $n > 0$ ، $\binom{-n}{r} = (-1)^r \binom{n+r-1}{r}$.

۵-۲۵. در بسط $(x+y+z)^n$ ، ضریب $x^3 y^2 z^3$ را بیابید.

۵-۲۶. در بسط $(2x+3y+4z+w)^9$ ، ضریب $x^3 y^2 z^3$ را بیابید.

۵-۲۷. با بیان تمام ضرایب دوجمله‌ای برحسب فاکتوریل‌ها و ساده کردن آنها به‌طور جبری، نشان دهید که:

$$\binom{n}{n_1, n_2, \dots, n_k} = \binom{n-1}{n_1-1, n_2, \dots, n_k} + \cdots + \binom{n-1}{n_1, n_2, \dots, n_k-1}.$$

خودآزمایی چهارگزینه‌ای

۵-۱. با اعداد ۱، ۲، ۳، ۴، ۵، چند عدد مضرب ۵، سه‌رقمی می‌توان ساخت به‌نحوی

که ارقامش تکراری نباشد؟

الف. ۳۶ ب. ۴۰ ج. ۲۰ د. ۱۶

۵-۲. یک کلاس احتمال دارای ۶ دانشجوی پسر و ۴ دانشجوی دختر است. پس از

برگزاری یک امتحان نمرات آنها مرتب می‌شوند. اگر مرتب کردن نمرات دانشجویان

پسر و دانشجویان دختر بین خودشان به‌طور جداگانه انجام گیرد، در این صورت چند حالت ممکن وجود دارد؟

الف. $10!$ ب. $2 \times 6! \times 4!$ ج. $6! \times 4!$ د. $\frac{6! \times 4!}{10!}$

۵-۳. در آزمون ۳۰ سؤالی چهارگزینه‌ای، به چند طریق می‌توان، با انتخاب تصادفی، به ۴ سؤال پاسخ درست داد؟

الف. $\frac{30!}{26!} \times 3^{26}$ ب. $\binom{30}{4} \times 3^{26}$ ج. $\binom{30}{4} \times 4^{26}$ د. $\frac{30!}{26!} \times 4^{26}$

۵-۴. به چند طریق می‌توان ۳ اسکناس ده‌هزار تومانی، ۳ اسکناس پنج‌هزار تومانی، و ۴ اسکناس دوهزار تومانی را بین ۱۰ نفر تقسیم کرد؟

الف. $\frac{10!}{3! \times 3! \times 4!}$ ب. $3! \times 3! \times 4!$ ج. $\binom{10}{3}$ د. $3 \times 3 \times 4$

۵-۵. معادله‌ی $x_1 + x_2 = 15$ چند جواب با اعداد صحیح غیرمنفی دارد؟

الف. ۱۶ ب. $\frac{15 \times 14}{2}$ ج. 15×14 د. ۱۵

۵-۶. کدام گزینه درست است؟

الف. $\binom{n}{r} = \binom{n-1}{r} + \binom{n-1}{r-1}$, $r = 1, 2, \dots, (n-1)$

ب. $\binom{n}{r} = \sum_{i=1}^{r+1} \binom{n-i}{r-1}$

ج. $\binom{3n}{n} = \sum_{i=0}^n \binom{n+i}{r}^r$

د. $a^n = \sum_{i=0}^n \binom{n}{r} (a+1)^r$

۵-۷. ضریب x^7 در بسط $\left(x + 2 + \frac{1}{x}\right)^8$ کدام گزینه است؟

الف. ۱۲۰ ب. ۱۱۲ ج. ۲۸ د. ۳۶

۵-۸. مقدار $\sum_{i=1}^n i \binom{n}{i}$ برابر کدام گزینه است؟

الف. 2^n ب. 2^8 ج. 2^{10} د. 2^{11}

۵-۹. مقدار $\binom{-3}{2}$ برابر کدام گزینه است؟

الف. $\binom{4}{2}$ ب. ۳ ج. -۱۵ د. $\binom{5}{3}$

۵-۱۰. مقدار $\binom{-5}{3}$ برابر کدام گزینه است؟

الف. $-\binom{7}{4}$ ب. $\binom{9}{3}$ ج. $-\binom{9}{6}$ د. $\binom{7}{3}$

۵-۱۱. معادله‌ی $x_1 + x_2 = b$, $b \in \mathbb{N}$ چند جواب با اعداد صحیح دارد؟

الف. b ب. $\binom{b}{2}$ ج. $b \times (b-1)$ د. $b+1$

۵-۱۲. به چند طریق می‌توان a کارت شارژ ده‌هزار تومانی، b کارت شارژ پنج‌هزار تومانی، و c کارت شارژ دوهزار تومانی را بین k نفر تقسیم کرد؟ ($k = a + b + c$)

الف. $\frac{k!}{a! \times b! \times c!}$ ب. $a! \times b! \times c!$ ج. $\binom{k}{3}$ د. $a \times b \times c$

۵-۱۳. حاصل عبارت $\frac{{}_nP_r}{{}_{n+1}P_{r+1}}$ کدام است؟

الف. $\frac{1}{n+1}$ ب. $\frac{r}{n}$ ج. $\frac{1}{(n+1)!}$ د. $\frac{r+1}{n+1}$

۵-۱۴. یک تاس و یک سکه را با هم پرتاب می‌کنیم. تعداد صورت‌هایی که در آن عدد زوج آمده، کدام است؟

الف. ۲ ب. ۳ ج. ۶ د. ۱

۵-۱۵. اگر ترکیب $\binom{a+b}{a} = m$ باشد، مقدار ترکیب $\binom{a+b}{b}$ کدام است؟

الف. m ب. bm ج. am د. $(a+m)m$

۵-۱۶. مجموعه‌ی $\{a, b, c, d, e\}$ چند زیرمجموعه‌ی دوعضوی دارد؟

الف. ۲۵ ب. ۲۰ ج. ۱۵ د. ۱۰

۵-۱۷. اگر ${}_nP_r = 20$ باشد، مقدار n کدام است؟

الف. ۲ ب. ۵ ج. ۱۰ د. ۲۰

۵-۱۸. با شرط $x \in \{1, 2, \dots\}$ ، جواب معادله‌ی $C_x^2 = 2x$ کدام است؟

الف. ۲. ب. ۳. ج. ۴. د. ۵.

۵-۱۹. تعداد جواب‌های صحیح و نامنفی نامعادله‌ی $x_1 + x_2 + x_3 \leq 7$ کدام است؟

الف. $\binom{7}{3}$. ب. $\binom{10}{3}$. ج. $\binom{9}{2}$. د. $\binom{11}{4}$.

۵-۲۰. حاصل عبارت $\binom{n+2}{r-1} + \binom{n+2}{r}$ کدام است؟

الف. $\binom{n+1}{r}$. ب. $\binom{n+3}{r-1}$. ج. $\binom{n+3}{r}$. د. $\binom{n+2}{r-1}$.

۵-۲۱. عدد طبیعی N چند مقسوم علیه دارد؟ ($N = P_1^{n_1} P_2^{n_2} \dots P_k^{n_k}$) که در آن

$P_1, \dots, P_k$ اعداد اول متمایزند)

الف. $\prod_{i=1}^k n_i$. ب. $\prod_{i=1}^k (1 + n_i)$. ج. $\prod_{i=1}^k (n_i - 1)$. د. $\prod_{i=1}^k (n_i + 1)$.

فصل ششم

احتمال

هدف کلی

یادگیری اصول موضوع احتمال و محاسبه‌ی احتمال پیشامدها تحت مدل هم‌شانسی یا مدل همگن

هدف‌های یادگیری

- خواننده پس از یادگیری این فصل باید بتواند:
- تعریف یک آزمایش تصادفی را بیان کند.
- مفاهیم و انواع فضای نمونه‌ای و پیشامد، و مفهوم برآمد را شرح دهد.
- سه اصل موضوع احتمال را بیان نماید.
- مدل هم‌شانسی برای فضای نمونه‌ای با تعداد عناصر متناهی را شرح دهد.
- قضایای حاصله از سه اصول موضوع احتمال را اثبات نماید.
- محاسبه‌ی اجتماع و اشتراک پیشامدها را انجام دهد.
- تعریف احتمال شرطی را بیان نماید و ویژگی‌های آن را اثبات کند.
- تعریف پیشامدهای مستقل را بیان نماید و برای آن مثال بیاورید.
- قضیه‌ی بیز را اثبات و مسائل آن را حل کند.
- در آزمایش‌های چندمرحله‌ای، از نمودار درختی برای محاسبه‌ی احتمال پیشامدها استفاده کند.

مقدمه

احتمال بیانی با یک مقدار عددی از درجه‌ی عدم قطعیت در مورد وقوع یک رویداد است. نظریه‌ی احتمال چارچوب تحلیلی برای علم آمار ایجاد می‌کند و به ما اجازه می‌دهد با استنتاج روی ترکیب نمونه‌ها مقادیر مورد انتظار کمی‌سازی شده را تعیین کنیم. قدمت تئوری احتمال جدید به بیش از هشتاد سال قبل برمی‌گردد؛ وقتی که ریاضی‌دانان می‌خواستند برخی مسائل مرتبط با افزایش درآمد را حل کنند. از نظر تاریخی، اولین بار محاسبه‌ی احتمال پیشامد در قرن شانزدهم میلادی توسط کاردانو^۱ (۱۵۷۶-۱۵۰۱) و گالیله^۲ (۱۶۴۲-۱۵۶۴) شروع شد. این دو توانستند احتمال ترکیب‌هایی از حالات مختلف پرتاب تاس را به صورت عددی محاسبه کنند. در این ارتباط پاسکال^۳ (۱۶۶۲-۱۶۲۳) و فرما^۴ (۱۶۶۵-۱۶۰۱) موفق شدند احتمال برد در بازی‌های معینی مربوط به قمار با تاس را محاسبه کنند. اما به طور اصولی، اساس و پایه پیشرفت این علم پس از تعریف اصولی احتمال توسط کلموگروف^۵ (۱۹۷۸-۱۹۰۳) ریاضیدان مشهور روسی قرن بیستم آغاز شد.

در این فصل سعی می‌شود به اختصار مفاهیم مقدماتی احتمال شامل پیشامد، چگونگی تخصیص احتمال به آنها، روابط پیشامدها، قواعدی از احتمال، احتمال شرطی، و قضیه‌ی بیز بیان شده و مثال‌هایی ارائه می‌شود.

۶-۱ آزمایش تصادفی

در بیشتر شاخه‌های علوم، آزمایش‌ها ابزارهای حیاتی هستند. آزمایش‌ها به دو گروه با نتیجه‌ی قطعی و نتیجه‌ی تصادفی تقسیم می‌شوند. در آمار و احتمال با آزمایش‌های تصادفی سروکار داریم که قبل از انجام نمی‌توانیم نتایج را با یقین کامل پیش‌بینی کنیم ولی کلیه‌ی نتایجی که بعد از انجام آن عمل ممکن است رخ دهد، قابل پیش‌بینی

^۱ Girolamo Cardano

^۲ Galileo Galilei

^۳ Blaise Pascal

^۴ Pierre Fermat

^۵ A. N. Kolmogorov

هستند.

- در بین ۱۰۰۰۰ پیچ همانند که با ماشینی ساخته شده‌اند، نمی‌توانیم تعداد پیچ‌های سالم یا معیوب را از قبل دقیقاً معین کنیم.

- با اطمینان نمی‌توانیم بگوییم در پرتاب تاس همگن کدام یک از ارقام ۱ تا ۶ ظاهر می‌شوند.

- در ظرفی که محتوی ۱۰ مهره‌ی همانند به شماره‌های ۱ تا ۱۰ است، نمی‌توان نتیجه‌ی انتخاب مهره را از قبل مشخص کرد.

آزمایش‌هایی که در بالا درباره‌ی آنها صحبت کردیم، دارای ویژگی‌های مشترک خاصی هستند. برای هر آزمایش، از قبل تمام برآمدهای ممکن را نمی‌دانیم و به این خاطر است که بعد از انجام هر آزمایش، هیچ برآمد رخ داده‌ای شگفت‌انگیز نیست. اما با وجود این، در حین انجام آزمایش نمی‌دانیم که برآمد معین چه خواهد بود یا به عبارت دیگر در زمان انجام آزمایش یک عدم اطمینان درباره‌ی نتیجه وجود دارد. علاوه بر این، آزمایش می‌تواند در شرایطی یکسان، تکرار شود. این ویژگی‌ها شرح یک آزمایش تصادفی (یا آماری) هستند.

تعریف ۶-۱. یک آزمایش تصادفی آزمایشی است که در آن:

الف. تمام برآمدهای ممکن آزمایش از قبل مشخص هستند،

ب. هر بار انجام آزمایش، منتج به نتیجه‌ای می‌شود که از قبل قطعی نیست،

ج. آزمایش می‌تواند در شرایطی یکسان تکرار شود.

از این به بعد مراد ما از آزمایش، همان آزمایش تصادفی است و به هر فرایند مشاهده یا اندازه‌گیری یا شمارش، آزمایش می‌گوییم.

۶-۲ فضای نمونه‌ای و برآمد

نتایجی که از یک آزمایش به دست می‌آیند، برآمدهای آزمایش نامیده می‌شوند.

مجموعه‌ی تمام برآمدهای ممکن آزمایش را **فضای نمونه‌ای** می‌خوانند و معمولاً آن را

با حرف S نشان می‌دهند. در فضای نمونه‌ای، هر برآمد را یک **عنصر فضای نمونه‌ای**

یا صرفاً یک **نقطه‌ی نمونه‌ای** می‌نامند. اگر فضای نمونه‌ای دارای تعداد متناهی از

عناصر باشد، می‌توان این عناصر را با نمادهای رایج در مجموعه‌ها، فهرست کرد.

مثال ۶-۱. فضای نمونه‌ای برای برآمدهای ممکن پرتاب یک سکه را می‌توان به صورت $S = \{H, T\}$ نوشت، که در آن H و T به ترتیب نمایش شیر و خط را می‌دهند.

مجموعه‌ی حاصل از برآمدها، یعنی S ، ممکن است متناهی، نامتناهی شمارا یا نامتناهی باشد.

مثال ۶-۲. (فضای نمونه‌ای متناهی) فضای نمونه‌ای از آزمایش پرتاب یک سکه و یک تاس عبارت است از:

$$S = \{(1, H), (1, T), (2, H), \dots, (6, T)\}.$$

این مجموعه ۱۲ عضو دارد، پس یک فضای نمونه‌ای متناهی است.

مثال ۶-۳. (فضای نمونه‌ای نامتناهی شمارا) سکه‌ای را آن‌قدر پرتاب می‌کنیم تا برای اولین بار شیر مشاهده شود. فضای نمونه‌ای این آزمایش دارای تعدادی متناهی عنصر نیست.

$$S = \{H, TH, TTH, \dots\}.$$

مثال ۶-۳. یک فضای نمونه‌ای نامتناهی شمارا که اعضای آن را می‌توان با مجموعه‌ی اعداد طبیعی متناظر نمود.

مثال ۶-۴. (فضای نمونه‌ای نامتناهی) مدت‌زمانی که دانشجویان رشته‌ی آمار و کاربردها، در طول یک روز از گوشی همراه‌شان استفاده می‌کنند، می‌تواند بازه‌ی $[0, 24]$ برحسب ساعت باشد. بنابراین $S = [0, 24]$ یک فضای نمونه‌ای نامتناهی است.

فضای نمونه‌ای به دو حالت گسسته و پیوسته، تقسیم‌بندی می‌شود. اگر مجموعه‌ی برآمدهای ممکنه از آزمایش، شمارا (متناهی یا نامتناهی شمارش‌پذیر) باشد، فضای نمونه‌ای را گسسته می‌گویند. در حالتی دیگر، اگر زیرفاصله‌ای از اعداد حقیقی، و یا

نامتناهی شمارش ناپذیر باشد، فضای نمونه‌ای، پیوسته می‌شود. مثال‌های ۶-۱، ۶-۲، و ۶-۳ فضای نمونه‌ای گسسته و مثال ۶-۴ فضای نمونه‌ای پیوسته هستند.

۶-۳ پیشامد

یک **پیشامد** زیرمجموعه‌ای از فضای نمونه‌ای است و معمولاً آن را با حروف $A, B, C, \dots$ نشان می‌دهیم. هرگاه برآمد آزمایش تصادفی عضوی از A باشد، گوییم پیشامد A رخ داده است.

مثال ۶-۵. در پرتاب یک تاس، $S = \{1, 2, 3, 4, 5, 6\}$ ، پیشامد A عبارت است از مشاهده‌ی عددی فرد. بنابراین A دارای عناصر ۱، ۳، و ۵ است.

$$A = \{1, 3, 5\} \subset S.$$

هر برآمدی مانند ۱ یا ۳ یا ۵ منجر به رخ دادن A می‌شود. پیشامد B که نتیجه‌ی پرتاب تاس عددی بخش‌پذیر بر ۳ است به صورت $B = \{3, 6\}$ است.

مثال ۶-۶. صفحه‌ی ساعتی به صورت دایره است. اگر بر اثر تصادف تیری به سمت صفحه‌ی ساعت پرتاب شود، اصابت تیر به ناحیه‌ی بین دو عقربه که ساعت ۱۲:۱۵ را تعیین می‌کند یک پیشامد محسوب می‌شود.

مثال ۶-۷. در میان بسته‌های کبریت پیشامد آنکه تعداد نخ‌های یک قوطی کبریت کمتر از ۳۰ باشد.

اگر پیشامد A تنها یک عنصر داشته باشد، آن را پیشامد ساده می‌نامند.

مثال ۶-۸. در پرتاب یک تاس، آمدن عدد ۶ یک پیشامد ساده است.

$$A = \{6\}.$$

اگر مجموعه‌ی A که زیرمجموعه‌ای از فضای نمونه‌ای است، تهی باشد، آن را **پیشامد محال** می‌نامند و با ϕ نشان می‌دهند و اگر A برابر S باشد آن را **پیشامد حتمی** می‌نامند.

مثال ۶-۹. در پرتاب یک تاس، مشاهده‌ی عدد ۲٫۵ یک پیشامد محال و پیشامد ظاهرشدن یک عدد طبیعی کوچکتر از ۷ یک پیشامد حتمی است.

اگر هر عضو که در A است، در B نیز باشد، می‌گویند A زیرپیشامد B است و می‌نویسند $A \subset B$.

مثال ۶-۱۰. خودروهایی را که در یک بزرگراه در حال حرکت هستند در نظر بگیرید. ملاحظه می‌کنید که A زیرپیشامد B است.

$$A = \{ \text{خودروهایی که سرعت آنها حداقل ۶۰ کیلومتر بر ساعت} \}$$

$$B = \{ \text{خودروهایی که سرعت آنها حداقل ۴۰ کیلومتر بر ساعت} \}$$

اگر A زیرپیشامد B و B زیرپیشامد A باشد، می‌گویند A و B پیشامدهای برابرند و می‌نویسند $A = B$.

مثال ۶-۱۱. در کتابخانه‌ای در طبقه‌ی دوم قفسه‌ایی به کتاب‌های تخصصی رشته‌ی آمار اختصاص داده‌اند. دو پیشامد زیر معادلند:

$$A = \{ \text{کتاب طبقه دوم قفسه} \},$$

$$B = \{ \text{کتاب تخصصی رشته‌ی آمار} \}.$$

اجتماع دو پیشامد، از تمام عنصرهای A یا B یا هر دو تشکیل می‌شود و آن را با $A \cup B$ نشان داده و پیشامد « A » یا « B » می‌نامند. این پیشامد موقعی رخ می‌دهد که لااقل یکی از دو پیشامد A و B رخ دهند.

مثال ۶-۱۲. در آزمایش تصادفی از پرتاب دو تاس، اگر A پیشامد مجموع دو پرتاب عدد هفت،

$$A = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}$$

و B پیشامد نتیجه‌ی یکی از تاس‌ها وجه سه باشد،

$$B = \{(3,1), (3,2), (3,3), (3,4), (3,5), (3,6), (1,3), (2,3), (4,3), (5,3), (6,3)\}$$

پیشامد اجتماع این دو عبارت است از:

$$A \cup B = \{(1,6), (2,5), (5,2), (6,1), (3,1), (3,2), (3,3), (3,4), (3,5), (3,6), (1,3), (2,3), (4,3), (5,3), (6,3)\}.$$

اشتراک پیشامد از تمام پیشامدهای ساده که در A و در B تشکیل می‌شود و آن را با $A \cap B$ نشان داده و پیشامد « A » و « B » می‌نامند. این پیشامد موقعی رخ می‌دهد که A و B هر دو با هم رخ دهند.

مثال ۶-۱۳. در مثال ۶-۱۲، ملاحظه می‌شود:

$$A \cap B = \{(3,4), (4,3)\}.$$

اگر A و B دو پیشامد از S باشند و $A \cap B = \emptyset$ آنگاه A و B را دو پیشامد «جدا» یا ناسازگار می‌نامند. در این حالت رخ دادن هر دو پیشامد با هم محال است و معمولاً می‌گویند A و B ناسازگارند.

مثال ۶-۱۴. در پرتاب سه سکه، پیشامدهای زیر همگی ناسازگارند.

$$A = \{HHH, HHT\}; B = \{HTH, THH, TTH\}; C = \{TTT, HTT, THT\}.$$

(آیا شرحی برای این سه پیشامد دارید؟)

تفاضل A از پیشامد B از تمام عناصر که در A بوده ولی در B نباشند تشکیل می‌شود، و آن را با $A - B$ نشان داده و پیشامد « A نه B » (یا تفاضل نسبی) می‌نامند. اگر $A \subset B$ باشد $A - B = \emptyset$ و $B - A$ را یک تفاضل واقعی می‌نامند زیرا B تمام A را دربردارد.

مثال ۶-۱۵. در پرتاب دو تاس، A پیشامد آنکه در بار دوم عدد شش مشاهده شود و B پیشامد آنکه در بار اول عدد شش مشاهده شود آنگاه:

$$A - B = \{(1,6), (2,6), (3,6), (4,6), (5,6)\},$$

به عبارت دیگر $A - B$ ، پیشامد آن است که فقط در بار دوم عدد شش ظاهر می‌شود. تذکر: توجه کنید که در حالت کلی، $A - B \neq B - A$ ، یعنی تفاضل نسبی ویژگی تقارن ندارد.

تفاضل پیشامد A از فضای نمونه، یعنی $S - A$ را متمم پیشامد A می‌گویند و آن را با A' نشان داده و پیشامد « A نه» می‌نامند. واضح است:

$$A \cup A' = S, \quad A \cap A' = \emptyset.$$

مثال ۶-۱۶. دو پیشامد زیر در پرتاب n بار یک سکه، متمم یکدیگرند:

A : پیشامد آنکه حداقل در یکی از پرتاب‌ها شیر مشاهده شود،

B : پیشامد آنکه نتیجه‌ی تمامی پرتاب‌ها خط باشد.

تفاضل متقارن دو پیشامد A و B عناصر آن یا عضوی از A ، یا عضوی از B ، ولی نه عضو هر دو تشکیل می‌شود و آن را با $A \Delta B$ نشان می‌دهند و پیشامد «یا A یا B » می‌نامند. واضح است که $B \Delta A = A \Delta B$ و از این رو کلمه‌ی متقارن را به کار می‌برند.

مثال ۶-۱۷. در مثال ۶-۱۵، دو پیشامد ذکر شده را در نظر بگیرید:

$$A \Delta B = \{(1,6), (2,6), (3,6), (4,6), (5,6), (6,5), (6,4), (6,3), (6,2), (6,1)\},$$

می‌توان از نتیجه‌ی این عملیات متوجه شد که مفهوم $A \Delta B$ یعنی هم‌زمان برای هر دو تاس عدد شش رخ نمی‌دهد.

۶-۴ احتمال

احتمال‌ها، مقادیر یک تابع مجموعه‌ای‌اند که اندازه‌ی احتمال هم خوانده می‌شوند، زیرا همان‌طور که خواهیم دید، این تابع، اعداد حقیقی در بازه‌ی $[0, 1]$ را به زیرمجموعه‌های مختلف فضای نمونه‌ای S نسبت می‌دهد. اصول موضوع احتمال که در اینجا فرمول‌بندی می‌کنیم، فقط وقتی به کار می‌روند که فضای نمونه‌ای S گسسته باشد. به عبارت دیگر فضای نمونه‌ای که تعداد عناصر آن متناهی، $S = \{O_1, O_2, O_3, \dots, O_n\}$ ، یا نامتناهی شمارش‌پذیر، $S = \{O_1, O_2, O_3, \dots\}$ ، باشد. علامت $P(A)$ برای نشان دادن احتمال پیشامد A به کار می‌رود. (اصول موضوع احتمال در حالت کلی، به مفهوم «سیگما-میدان» گره خورده است که در این سطح از درس ضرورتی به بیان آن نیست.)

اصل موضوع ۱. احتمال یک پیشامد، عددی حقیقی نامنفی است، یعنی برای هر زیرمجموعه‌ی A از S ، $P(A) \geq 0$.

اصل موضوع ۲. $P(S) = 1$.

اصل موضوع ۳. اگر $A_1, A_2, A_3, \dots$ دنباله‌ای متناهی یا نامتناهی از پیشامدهای دوبه‌دو

ناسازگار در S باشند، آنگاه:

$$P(A \cup A_+ \cup A_+ \cup \dots) = P(A) + P(A_+) + P(A_+) + \dots.$$

اصول موضوع ذاتاً نیازی به برهان ندارند، اما اگر بخواهیم نظریه‌ی حاصل را به کار ببریم، وقتی به احتمال‌ها یک مفهوم «حقیقی» می‌دهیم باید نشان دهیم که اصول موضوع صادق هستند.

به‌طورشهودی، ابتدا احتمال منسوب به یک پیشامد را به‌عنوان معیاری عددی در نظر می‌گیریم که با آن، میزان انتظار برای وقوع پیشامد را وقتی آزمایش اجرا می‌شود اندازه می‌گیرند. به منظور تعیین این میزان، طبیعی است که نسبت دفعات وقوع پیشامد در تکرار آزمایش را در نظر بگیریم. تصور شهودی ما از اندازه‌ی عددی احتمال یک پیشامد، نسبت دفعاتی است که انتظار می‌رود آن پیشامد وقتی که آزمایش در شرایطی یکسان تکرار شود، رخ بدهد.

فرایند مناسب برای تعیین احتمال پیشامدها، بستگی به طبیعت آزمایش و فضای نمونه‌ای مربوط است. در بعضی موارد، می‌توان نسبت دفعاتی را که انتظار می‌رود هر برآمد مقدماتی رخ بدهد، با استنتاج منطقی و بدون آنکه آزمایش عملاً اجرا شود، تعیین کرد. در موارد دیگر، لازم است آزمایش را برای تعداد زیادی از دفعات تکرار کرد تا اطلاعاتی در مورد فراوانی وقوع برآمدهای گوناگون به دست آید. تشریح این دو نوع حالت، ما را به‌سوی تعریف رسمی احتمال راهنمایی می‌کند.

چون نسبت‌ها همواره مثبت یا صفرند، اولین اصل موضوع با تعبیر فراوانی در هماهنگی کامل است. اصل موضوع دوم به‌طور غیرمستقیم بیان می‌کند که حتمیت با احتمال ۱، یکی است، روی هم‌رفته همیشه می‌پذیریم که باید یکی از امکانات موجود در S رخ دهد، و به این پیشامد که حتمی است احتمال ۱ را نسبت می‌دهیم. تا آنجا که به تعبیر فراوانی مربوط است، احتمال ۱ اشاره براین دارد که پیشامد مورد بحث در ۱۰۰ درصد مواقع رخ خواهد داد، یا به عبارت دیگر، این پیشامد مطمئناً رخ می‌دهد.

با در نظر گرفتن سومین اصل موضوع در ساده‌ترین حالت، یعنی برای دو پیشامد ناسازگار A و A_+ ، می‌توان به سادگی دید که این اصل موضوع از نظر تعبیر فراوانی برآورده می‌شود. اگر پیشامدی مثلاً در ۲۸ درصد مواقع، و پیشامد دیگری در ۳۹ درصد مواقع رخ دهد، و هر دو پیشامد نتوانند به طور هم‌زمان رخ دهند (یعنی، دوه‌دو

ناسازگار باشند)، آنگاه یکی یا دیگری در $28 + 39 = 67$ درصد مواقع رخ خواهد داد. بنابراین سومین اصل موضوع صادق است و وقتی بیش از دو پیشامد دویه‌دو ناسازگار وجود داشته باشند، همین نوع استدلال به کار می‌رود.

قبل از اینکه برخی نتایج فوری اصول موضوع احتمال را مطالعه کنیم، بر این نکته تأکید می‌کنیم که این سه اصل موضوع به ما نمی‌گویند چگونه احتمال‌ها را به پیشامدها تخصیص دهیم، بلکه فقط راه‌های انجام این کار را محدود می‌کنند.

مثال ۶-۱۸. در یک آزمایش، فضای نمونه‌ای ممکن به چهار پیشامد دویه‌دو ناسازگار A ، B ، C و D را افراز شده است. برای هریک از موارد زیر، توضیح دهید که چرا راهی مجاز برای تخصیص احتمال‌ها وجود ندارد.

الف. $P(A) = 0.12$ ، $P(B) = 0.63$ ، $P(C) = 0.45$ ، $P(D) = -0.20$

ب. $P(A) = \frac{9}{120}$ ، $P(B) = \frac{45}{120}$ ، $P(C) = \frac{27}{120}$ ، $P(D) = \frac{46}{120}$

حل. الف. $P(D) = -0.20$ ناقض اصل موضوع ۱ است.

ب. $P(S) = P(A \cup B \cup C \cup D) = \frac{9}{120} + \frac{45}{120} + \frac{27}{120} + \frac{46}{120} = \frac{127}{120} \neq 1$

به عبارت دیگر فضای نمونه‌ای مقدار احتمالی بزرگتر از یک دارد.

البته، در کاربرد واقعی، احتمال‌ها بر مبنای تجارب گذشته، بر مبنای تحلیل محتاطانه‌ی تمام شرایط زیربنایی، بر مبنای دآوری‌های ذهنی، یا بر مبنای مفروضات، گاهی با فرض هم‌شانس بودن (یا به معنای دیگر، برابری احتمال‌ها برای) تمام برآمدهای ممکن، نسبت داده می‌شوند.

در تخصیص احتمال به پیشامدها در یک فضای نمونه‌ای، ضروری نیست که احتمال هر زیرمجموعه‌ی ممکن را توصیف کنیم، زیرا یک فضای نمونه‌ای با فقط ۲۰ برآمد ممکن، $2^{20} = 1,048,576$ زیرمجموعه دارد و تعداد زیرمجموعه‌ها وقتی ۵۰ برآمد ممکن، ۱۰۰ برآمد ممکن، یا بیشتر موجود باشند، بسیار سریع افزایش می‌یابد. اغلب به جای فهرست کردن احتمال‌های تمام زیرمجموعه‌های ممکن، احتمال‌های پیشامدهای، یا عناصر فضای نمونه‌ای S را فهرست می‌کنیم، و آنگاه قضیه‌ی زیر را به کار می‌بریم.

قضیه ۱-۶. اگر A پیشامدی از فضای نمونه‌ای گسسته‌ی S باشد، آنگاه $P(A)$ برابر است با مجموع احتمال‌های پیشامدهای ساده‌ی تشکیل دهنده‌ی A .
برهان. فرض می‌کنیم $O_1, O_2, O_3, \dots$ دنباله‌ای متناهی یا نامتناهی از پیشامدهای ساده باشد که پیشامد A را تشکیل می‌دهند. پس:

$$A = O_1 \cup O_2 \cup O_3 \cup \dots$$

و چون پیشامدهای ساده، O_i ها، برحسب تعریف دوه‌دو ناسازگارند، از اصل سوم احتمال نتیجه می‌شود که:

$$P(A) = P(O_1) + P(O_2) + P(O_3) + \dots$$

و این رابطه، برهان را کامل می‌کند.

برای استفاده از این قضیه، باید قادر به تخصیص احتمال‌ها به پیشامدهای ساده آزمایش باشیم. اینکه چگونه می‌توان این کار را در برخی حالات‌های خاص انجام داد، با مثال‌های زیر تشریح می‌شود.

مثال ۶-۱۹. اگر سکه‌ی سالمی را دو بار پرتاب کنیم، احتمال به دست آوردن حداقل یک شیر چقدر است؟

حل. فضای نمونه‌ای برای این آزمایش عبارت است از:

$$S = \{HH, HT, TH, TT\}$$

که در آن، H و T به ترتیب شیر و خط را نشان می‌دهند. چون سکه متعادل است، می‌پذیریم که وقوع این برآمدها هم‌شانس‌اند، و بنابراین به هر نقطه‌ی فضای نمونه‌ای احتمال $\frac{1}{4}$ را نسبت می‌دهیم، اگر A پیشامدی باشد که حداقل یک شیر به دست آید، آنگاه $A = \{HH, HT, TH\}$ ، و

$$\begin{aligned} P(A) &= P(\{HH\}) + P(\{HT\}) + P(\{TH\}) \\ &= \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}. \end{aligned}$$

مثال ۶-۲۰. تاسی را که احتمال وقوع هر عدد فرد با آن، دوبرابر احتمال وقوع هر عدد زوج است در نظر بگیرید. اگر در یک‌بار ریختن این تاس، G پیشامد وقوع عددی بزرگتر از ۳ باشد، $P(G)$ را بیابید.

حل. فضای نمونه‌ای عبارت است از $S = \{1, 2, 3, 4, 5, 6\}$. اگر به هر عدد زوج، احتمال w و به هر عدد فرد، احتمال $2w$ را نسبت دهیم، مطابق اصل موضوع ۲ داریم:

$$2w + w + 2w + w + 2w + w = 9w = 1$$

پس $w = \frac{1}{9}$ ، و

$$P(G) = \frac{1}{9} + \frac{2}{9} + \frac{1}{9} = \frac{4}{9}.$$

اگر فضای نمونه‌ای، گسسته ولی نامتناهی باشد، احتمال‌ها را باید به وسیله‌ی یک قاعده‌ی ریاضی ترجیحاً به وسیله‌ی فرمول یا معادله‌ای، به برآمدهای فردی نسبت داد.

مثال ۶-۲۱. اگر $O_1, O_2, O_3, \dots$ دنباله‌ای نامتناهی از پیشامدهای ساده‌ی یک آزمایش مفروض باشد، تحقیق کنید که:

$$P(O_i) = \left(\frac{1}{2}\right)^i \quad i = 1, 2, 3, \dots,$$

واقعاً یک اندازه‌ی احتمال است.

حل. چون مقادیر احتمال‌ها همگی مثبت‌اند، فقط باید نشان دهیم که $P(S) = 1$ داریم:

$$P(S) = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \dots$$

اگر فرمول تعیین مجموع جملات یک تصاعد هندسی نامتناهی را به کار ببریم، به دست می‌آوریم:

$$P(S) = \frac{\frac{1}{2}}{1 - \frac{1}{2}} = 1.$$

در رابطه با مثال ۶-۲۱، کلمه‌ی «مجموع» در قضیه‌ی ۶-۱، باید به قسمی تعبیر شود که شامل مقدار یک سری نامتناهی نیز باشد.

اندازه‌ی احتمال مثال ۶-۲۱، مثلاً اگر O_i این پیشامد باشد که شخصی در پرتاب سکه‌ای همگن برای اولین بار در پرتاب i م سکه، خط بیاورد، اندازه‌ی احتمال مناسبی است. پس، احتمال آنکه اولین خط در سومین، چهارمین، یا پنجمین پرتاب سکه ظاهر شود برابر است با:

$$\left(\frac{1}{2}\right)^3 + \left(\frac{1}{2}\right)^4 + \left(\frac{1}{2}\right)^5 = \frac{7}{32}$$

و احتمال آنکه اولین خط در تعداد فردی از پرتاب‌ها رخ دهد برابر است با:

$$\left(\frac{1}{2}\right)^1 + \left(\frac{1}{2}\right)^3 + \left(\frac{1}{2}\right)^5 + \dots = \frac{\frac{1}{2}}{1 - \frac{1}{4}} = \frac{2}{3}.$$

اگر مانند مثال ۶-۱۹، آزمایش چنان باشد که بتوانیم برای نقاط فضای نمونه‌ای S ، احتمال‌هایی برابر در نظر گیریم، می‌توانیم از حالت خاص قضیه‌ی ۶-۱ که به صورت قضیه‌ی ۶-۲ است، استفاده کنیم.

قضیه‌ی ۶-۲. اگر نتیجه‌ی آزمایشی بتواند به یکی از N برآمد مختلف هم‌شانس باشد، و n تا از این برآمدها با هم پیشامد A را تشکیل دهند، آنگاه احتمال پیشامد A برابر است با:

$$P(A) = \frac{n}{N}.$$

برهان. فرض می‌کنیم $O_1, O_2, \dots, O_N$ ، پیشامدهای ساده، یا پیشامدهایی با اعضای برآمدهای فردی S را نشان دهند که احتمال هر کدام $\frac{1}{N}$ است. اگر پیشامد A ، اجتماع n تا از این برآمدهای دوبه‌دو ناسازگار بوده، و مهم نباشد که کدام یک از آنها این پیشامد را تشکیل دهند، آنگاه:

$$\begin{aligned} P(A) &= P(O_1 \cup O_2 \cup \dots \cup O_n) \\ &= P(O_1) + P(O_2) + \dots + P(O_n) \\ &= \underbrace{\frac{1}{N} + \frac{1}{N} + \dots + \frac{1}{N}}_{n \text{ بار}} \\ &= \frac{n}{N}. \end{aligned}$$

مشاهده کنید که فرمول $P(A) = \frac{n}{N}$ در قضیه‌ی ۶-۲، با مفهوم احتمال کلاسیک یکی است (در برخی منابع از قضیه‌ی ۶-۲ به عنوان تعریفی از مفهوم احتمال کلاسیک

استفاده می‌کنند (بازرگان، ۱۳۸۴). البته آنچه در اینجا نشان داده‌ایم آن است که مفهوم احتمال کلاسیک با اصول موضوع احتمال سازگار است - این مطلب در مورد خاصی که برآمدهای فردی هم‌شانسانند از اصول موضوع احتمال نتیجه می‌شود.

مثال ۶-۲۲. در یک بسته ۴۸ پیراهن وجود دارد (پیراهن‌ها در چهار رنگ و در اندازه‌های یک الی دوازده تولید شده‌اند)، پنج پیراهن به تصادف اختیار می‌کنیم، احتمال اینکه سه پیراهن دارای یک شماره و دو پیراهن دیگر هم دارای یک شماره باشند، چقدر است؟

حل. تعداد راه‌هایی که می‌توان سه پیراهن با یک اندازه و دو پیراهن با یک اندازه‌ی دیگر، مثلاً سه پیراهن با اندازه‌ی ۷ و دو پیراهن با اندازه‌ی ۱ انتخاب کرد، برابر است با $\binom{4}{2} \times \binom{4}{3}$. چون ۱۲ اندازه از هر رنگ داریم ۱۲ راه انتخاب سه پیراهن با یک اندازه و سپس ۱۱ راه انتخاب دو پیراهن با یک اندازه وجود دارند. کل راه‌های ممکن انتخاب چنین پنج پیراهنی برابر

$$n = 12 \times 11 \times \binom{4}{3} \times \binom{4}{2}$$

است. از طرفی تعداد راه‌های ممکن انتخاب پنج پیراهن از ۴۸ پیراهن

$$N = \binom{48}{5}$$

است و لذا از قضیه‌ی ۶-۲ نتیجه می‌شود که احتمال پیشامد مطلوب عبارت است از:

$$P(A) = \frac{n}{N} = \frac{12 \times 11 \times \binom{4}{3} \times \binom{4}{2}}{\binom{48}{5}} = 0.0006.$$

۶-۵ برخی قواعد احتمال

با استفاده از سه اصل موضوع احتمال، می‌توانیم بسیاری از قاعده‌های دیگر را که کاربردهای مهمی دارند، نتیجه بگیریم. از پیامدهای فوری اصول موضوع، قضایای زیر را ثابت می‌کنیم.

قضیه ۶-۳. همواره داریم $P(\phi) = 0$.

برهان. چون پیشامدهای S و ϕ ناسازگارند و مطابق تعریف مجموعه‌ی تهی ϕ ،
 $S \cup \phi = S$ ، نتیجه می‌شود که:

$$\begin{aligned} P(S) &= P(S \cup \phi) \\ &= P(S) + P(\phi) \end{aligned} \quad (\text{بنابه اصل موضوع ۳})$$

و در نتیجه $P(\phi) = 0$.

باید به این مطلب مهم توجه کرد که از $P(A) = 0$ نتیجه نمی‌شود که الزاماً A مجموعه‌ای تهی است. در عمل اغلب به پیشامدهایی که، به زبان محاوره‌ای، در یک میلیون سال هم رخ نمی‌دهند، احتمال صفر را نسبت می‌دهیم. برای نمونه، مثال کلاسیکی وجود دارد که ما به این پیشامد که میمونی «جمهوری افلاطون» را کلمه‌به‌کلمه بدون خطا با فشار روی دکمه‌ها، تایپ کند، احتمال صفر را نسبت می‌دهیم. این واقعیت شایسته‌ی توجه است که $P(A) = 0$ ، رابطه‌ی $A = \phi$ را، به ویژه در فضای نمونه‌ای ناشمارا، نتیجه نمی‌دهد.

قضیه ۶-۴. برای هر دو پیشامد ناسازگار از S مانند A و B داریم:

$$P(A \cup B) = P(A) + P(B).$$

برهان. با توجه به رابطه‌ی $A \cup B = A \cup B \cup \phi \cup \phi \cup \dots$ و اصل موضوع ۳،

$$\begin{aligned} P(A \cup B) &= P(A \cup B \cup \phi \cup \phi \cup \dots) \\ &= P(A) + P(B) + P(\phi) + P(\phi) + \dots \\ &= P(A) + P(B). \end{aligned}$$

یک نتیجه‌ی فوری که می‌توان از قضیه ۶-۴ گرفت آن است که از ناسازگاری A' و A و اینکه $S = A \cup A'$ ، استفاده کنیم و رابطه‌ی زیر را به دست آوریم:

$$P(A') = 1 - P(A).$$

در رابطه با تعبیر فراوانی، این نتیجه اشاره بر آن دارد که اگر پیشامدی، مثلاً در ۳۷ درصد مواقع رخ دهد، در ۶۳ درصد مواقع رخ نمی‌دهد.

قضیه ۵-۶. اگر A و B پیشامدهایی در فضای نمونه‌ای S باشند و $A \subset B$ ، آنگاه $P(A) \leq P(B)$.

برهان. چون $A \subset B$ ، می‌توانیم بنویسیم:

$$B = B \cap S = B \cap (A \cup A') = (B \cap A) \cup (B \cap A') = A \cup (B \cap A')$$

که درستی روابط آخر به‌آسانی به وسیله‌ی یک نمودار ون تحقیق می‌شود، در این صورت A و $A' \cap B$ ناسازگارند، و در نتیجه:

$$\begin{aligned} P(B) &= P(A) + P(A' \cap B) && \text{(بنابه اصل موضوع ۳)} \\ &\geq P(A) && \text{(بنابه اصل موضوع ۱)} \end{aligned}$$

با استفاده از کلمات، این قضیه بیان می‌کند که اگر پیشامد A زیرمجموعه‌ای از پیشامد B باشد، آنگاه $P(A)$ بزرگتر از $P(B)$ نیست. مثلاً، احتمال کشیدن یک کارت سیاه از یک دسته کارت ۵۲ تایی، بزرگتر از احتمال کشیدن یک کارت سیاه یا قرمز نیست. در واقع احتمال $\frac{1}{4}$ با احتمال $\frac{1}{4}$ مقایسه می‌شود.

قضیه ۶-۶. برای هر پیشامد A ، $0 \leq P(A) \leq 1$.

برهان. با به‌کاربردن قضیه ۵-۶ و این واقعیت که برای هر پیشامد A در S ، $\phi \subset A \subset S$ داریم:

$$P(\phi) \leq P(A) \leq P(S)$$

چون $P(\phi) = 0$ و $P(S) = 1$ خواهیم داشت:

$$0 \leq P(A) \leq 1.$$

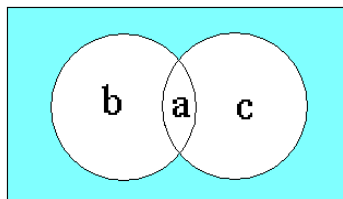
سومین اصل موضوع احتمال اغلب به یک **قاعده‌ی جمع خاص** اطلاق می‌شود. قاعده‌ی مزبور از این نظر خاص است که پیشامدهای $A_1, A_2, A_3, \dots$ باید دوه‌دو ناسازگار باشند. برای هر دو پیشامد A و B ، **قاعده‌ی جمع عام** در قضیه ۶-۷ موجود است.

قضیه ۶-۷. اگر A و B دو پیشامد در فضای نمونه‌ای S باشند، آنگاه:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

برهان. همان‌طور که در نمودار ون شکل ۶-۱ دیده می‌شود به پیشامدهای دوه‌دو ناسازگار $A \cap B$ ، $A \cap B'$ ، $A' \cap B$ ، به ترتیب احتمال‌های a ، b ، c را نسبت می‌دهیم و به دست می‌آوریم:

$$\begin{aligned} P(A \cup B) &= a + b + c \\ &= (a + b) + (c + a) - a \\ &= P(A) + P(B) - P(A \cap B). \end{aligned}$$



شکل ۶-۱. نمودار ون برای اثبات قضیه‌ی ۶-۷

مثال ۶-۲۳. اگر برای خانواده‌ای (که به منظور یک بررسی نمونه‌ای در یک ناحیه‌ی بزرگ شهری، به تصادف انتخاب شده است) احتمال دارا بودن یک دستگاه تلویزیون رنگی، یک دستگاه کامپیوتر و یا هر دو وسیله به ترتیب $۰/۸۶$ ، $۰/۳۵$ و $۰/۲۹۴$ باشد، احتمال اینکه این خانواده یکی از دو نوع یا هر دو نوع دستگاه (حداقل یکی از دو نوع دستگاه) را داشته باشند چقدر است؟

حل. اگر A ، این پیشامد باشد که خانواده‌ی مزبور دارای یک دستگاه تلویزیون رنگی است، و B ، این پیشامد باشد که دارای یک دستگاه رایانه است، داریم $P(A \cap B) = ۰/۲۹$ ، $P(A) = ۰/۸۶$ ، $P(B) = ۰/۳۵$ و اگر این مقادیر را در فرمول قضیه‌ی ۶-۷ قرار دهیم، نتیجه می‌شود:

$$P(A \cup B) = ۰/۸۶ + ۰/۳۵ - ۰/۲۹ = ۰/۹۲.$$

مثال ۶-۲۴. خودرویی که در کنار بزرگ‌راهی متوقف شده است با احتمال $۰/۲۳$ ترمزهای معیوب و با احتمال $۰/۲۴$ فرسودگی شدید تایر دارد. همچنین با احتمال $۰/۳۸$ ترمزهای معیوب یا فرسودگی شدید تایر یا هر دو را دارد. احتمال اینکه این اتومبیل، ترمزهایش معیوب بوده و فرسودگی شدید تایر داشته باشد، چقدر است؟

حل. اگر B این پیشامد باشد که اتومبیل مزبور ترمزهایش معیوب است، و T این پیشامد باشد که تایرهايش فرسودگی شدید دارند، داریم:

$$P(B) = 0.23, \quad P(T) = 0.24, \quad P(B \cup T) = 0.38.$$

از قرارداد این مقادیر در فرمول قضیه ی ۶-۷ نتیجه می شود:

$$0.38 = 0.23 + 0.24 - P(B \cap T).$$

از حل این معادله نسبت به $P(B \cap T)$ ، به دست می آوریم:

$$P(B \cap T) = 0.23 + 0.24 - 0.38 = 0.09.$$

با به کار بردن مکرر قضیه ی ۶-۷، می توان قاعده ی جمع را به قسمی تعمیم داد که برای هر تعداد از پیشامدها قابل اجرا باشد. مثلاً برای سه پیشامد به دست می آوریم:

قضیه ی ۶-۸ اگر A ، B ، و C ، سه پیشامد دلخواه در فضای نمونه ای در S باشند، آنگاه:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C).$$

برهان. $A \cup B \cup C$ را به صورت $A \cup (B \cup C)$ می نویسیم و دوبار قضیه ی ۶-۷ را به کار می بریم، یک بار برای $P(A \cup (B \cup C))$ و یک بار برای $P(B \cup C)$ ، به دست می آوریم:

$$\begin{aligned} P(A \cup B \cup C) &= P(A \cup (B \cup C)) \\ &= P(A) + P(B \cup C) - P(A \cap (B \cup C)) \\ &= P(A) + P(B) + P(C) - P(B \cap C) - P(A \cap (B \cup C)). \end{aligned}$$

از قانون توزیع پذیری اشتراک نسبت به اجتماع استفاده نموده و داریم:

$$\begin{aligned} P(A \cap (B \cup C)) &= P((A \cap B) \cup (A \cap C)) \\ &= P(A \cap B) + P(A \cap C) - P((A \cap B) \cap (A \cap C)) \\ &= P(A \cap B) + P(A \cap C) - P(A \cap B \cap C), \end{aligned}$$

و نتیجه آنکه:

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) \\ &\quad - P(B \cap C) + P(A \cap B \cap C). \end{aligned}$$

مثال ۶-۲۵. فرض می‌کنیم شخصی که به دندان‌پزشک خود مراجعه می‌کند، احتمال اینکه دندان‌هایش را جرم‌گیری کند 0.44 ، احتمال اینکه دندانانی برای پرکردن داشته باشد، 0.24 ، احتمال آنکه دندانانی برای کشیدن داشته باشد 0.21 ، احتمال آنکه دندان‌هایش را جرم‌گیری و دندانانی را پرکند، 0.08 ، احتمال اینکه دندان‌هایش را جرم‌گیری کند و دندانانی را بکشد 0.11 ، احتمال اینکه دندانانی برای پرکردن و دندانانی برای کشیدن داشته باشد، 0.07 و احتمال اینکه دندان‌هایش را جرم‌گیری و دندانانی را پرکند و دندانانی را بکشد 0.03 باشد. احتمال اینکه دندان‌پزشک حداقل یکی از سه مورد را برای او انجام دهد چقدر است؟

حل. اگر C این پیشامد باشد که شخص دندان‌هایش را جرم‌گیری کند،

F این پیشامد باشد که شخص دندانانی را پرکند،

E این پیشامد باشد که شخص دندانانی را بکشد،

آنگاه $P(C) = 0.44$ ، $P(F) = 0.24$ ، $P(E) = 0.21$ ، $P(C \cap F) = 0.08$ ،

$P(C \cap E) = 0.11$ ، $P(F \cap E) = 0.07$ ، و $P(C \cap F \cap E) = 0.03$ ، و با قراردادن در

فرمول قضیه‌ی ۶-۸، نتیجه می‌شود که:

$$P(C \cup F \cup E) = 0.44 + 0.24 + 0.21 - 0.08 - 0.11 - 0.07 + 0.03 = 0.66.$$

۶-۶ احتمال شرطی

در اکثر مواقع لازم می‌شود که احتمال پیشامدی مانند A ، که با پیشامدی مانند B مربوط باشد، بعد از اطلاع از وقوع یا عدم وقوع پیشامد B ، اصلاح شود. به بیانی دیگر، ممکن است وقوع پیشامد B بر احتمال وقوع پیشامدی چون A تاثیر داشته باشد. بنابراین کسب اطلاعات درباره‌ی جنبه‌ای از نتایج آزمایش، ممکن است تجدیدنظر در احتمال پیشامدی را که مربوط به جنبه‌ی دیگری از نتایج است، ایجاب کند. احتمال تجدیدنظرشده‌ی A ، وقتی معلوم شود که B رخ داده است، **احتمال شرطی** A به شرط B نامیده و با $P(A|B)$ نشان داده می‌شود.

برخی از ایده‌هایی که به احتمال‌های شرطی مربوط می‌شوند، در مثال زیر تشریح شده‌اند.

مثال ۶-۲۶. در یک تحقیق از عملکرد بانکی، پنجاه شعبه مورد مطالعه قرار گرفته‌اند و

نتایج در جدول ۶-۱ خلاصه شده‌اند.

جدول ۶-۱. اطلاعات مربوط به مثال ۶-۲۶

		سرویس‌دهی به مشتریان	
		خوب	ضعیف
زمان گشایش شعبه	حداقل ۱۰ سال	۱۶	۴
	کمتر از ۱۰ سال	۱۰	۲۰

اگر فردی به تصادف یکی از این شعب را انتخاب کند، احتمال اینکه شعبه سرویس‌دهی خوبی به مشتریان ارائه دهد چقدر است؟ همچنین اگر شخصی به تصادف یکی از شعب را انتخاب کند که حداقل بیش از ده سال سابقه داشته باشد، احتمال اینکه شعبه انتخابی به مشتریان خوب سرویس‌دهی کند چقدر است؟ منظور از «تصادف» این است که در هر حالت، تمام انتخاب‌های ممکن هم‌شانس‌اند.

حل. با فرض هم‌شانسی همه‌ی مشتریان، فرمول قضیه‌ی ۶-۲ را به کار می‌بریم. اگر G ، انتخاب شعبه را نشان دهد که سرویس خوب ارائه می‌دهد و اگر فرض کنیم $n(G)$ تعداد اعضای G را نشان دهد، و $n(S)$ تعداد اعضای فضای نمونه‌ای باشد، به دست می‌آوریم:

$$P(G) = \frac{n(G)}{n(S)} = \frac{۱۶+۱۰}{۵۰} = ۰٫۵۲.$$

این پاسخ سؤال اول است.

برای پاسخ سؤال دوم، ما خود را به فضای نمونه‌ای کوچکتری که شامل سطر اول جدول است، یعنی به $۱۶+۴=۲۰$ شعب که حداقل ده سال سابقه دارند محدود می‌کنیم. از ۲۰ شعبه، ۱۶ شعبه سرویس خوب ارائه می‌دهند، و به دست می‌آوریم:

$$P(G|T) = \frac{۱۶}{۲۰} = ۰٫۸۰،$$

که در آن T ، انتخاب شعبه‌ای را نشان می‌دهد که حداقل ده سال سابقه دارد. این پاسخ سؤال دوم است و همان‌طور که انتظار داشتیم، $P(G|T)$ خیلی از $P(G)$ بزرگتر است.

چون صورت $P(G|T)$ برابر $n(T \cap G) = ۱۶$ ، یعنی تعداد شعبی است که حداقل ده سال سابقه دارند و سرویس‌دهی خوبی دارند، و مخرج برابر $n(T)$ ، یعنی

شعبی است که حداقل ده سال سابقه‌ی کار دارند. در واقع می‌توانیم آن را به صورت نمادی بنویسیم:

$$P(G|T) = \frac{n(T \cap G)}{n(T)}.$$

پس، اگر صورت و مخرج را به $n(S)$ ، تعداد کل شعب موردنظر، تقسیم کنیم، به دست می‌آوریم:

$$P(G|T) = \frac{\frac{n(T \cap G)}{n(S)}}{\frac{n(T)}{n(S)}} = \frac{P(T \cap G)}{P(T)}$$

و لذا احتمال شرطی $P(G|T)$ را برحسب دو احتمالی که برای تمام فضای نمونه‌ای S تعریف شده‌اند بیان کرده‌ایم.

با تعمیم این مثال، اکنون تعریف احتمال شرطی را به صورت زیر ارائه می‌دهیم.

تعریف ۶-۲. اگر A و B دو پیشامد دلخواه در فضای نمونه‌ای S باشند و $P(A) \neq 0$ ، احتمال شرطی B به شرط A به صورت زیر تعریف می‌شود:

$$P(B|A) = \frac{P(A \cap B)}{P(A)}.$$

مثال ۶-۲۷. با مراجعه به مثال ۶-۲۶، احتمال اینکه یکی از شعب با سابقه‌ی کمتر از ده سال، سرویس خوبی به مشتریان ارائه دهد چقدر است؟

حل. چون $P(T' \cap G) = \frac{10}{50} = 0.20$ و $P(T') = \frac{10+20}{50} = 0.60$ ، اگر مقادیر را در فرمول قرار دهیم نتیجه می‌دهد که:

$$P(G|T') = \frac{P(T' \cap G)}{P(T')} = \frac{0.20}{0.60}.$$

گرچه ما $P(B|A)$ را با مثالی که در آن تمام حالت‌ها هم‌شانس بودند معرفی کردیم، اما این مطلب، شرطی برای به کار بردن تعریف نیست.

مثال ۶-۲۸. با رجوع به تاس ناهمگن مثال ۶-۲۰، احتمال اینکه شماره‌ای که ظاهر می‌شود مربع کامل باشد چقدر است؟ همچنین به فرض اینکه این شماره بزرگتر از ۳ باشد، احتمال اینکه مربع کامل باشد چقدر است؟

حل. اگر A این پیشامد باشد که شماره‌ی ظاهرشده بزرگتر از ۳ است، و B این پیشامد باشد که شماره‌ی ظاهرشده مربع کامل است، داریم $A = \{4, 5, 6\}$ ، $B = \{1, 4\}$ ، و $A \cap B = \{4\}$. چون احتمال‌های ظاهرشدن ۱، ۲، ۳، ۴، ۵، ۶ با تاس موردنظر

به ترتیب $\frac{1}{9}$ ، $\frac{2}{9}$ ، $\frac{1}{9}$ ، $\frac{2}{9}$ ، $\frac{1}{9}$ ، $\frac{2}{9}$ هستند، پاسخ اولین سوال را به صورت

$$P(B) = \frac{2}{9} + \frac{1}{9} = \frac{1}{3}$$

پیدا می‌کنیم. برای تعیین $P(B|A)$ ابتدا

$$P(A) = \frac{1}{9} + \frac{2}{9} + \frac{1}{9} = \frac{4}{9} \quad \text{و} \quad P(A \cap B) = \frac{1}{9}$$

را محاسبه می‌کنیم. لذا با قراردادن این دو مقدار در فرمول تعریف ۶-۲، به دست می‌آوریم:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{1}{9}}{\frac{4}{9}} = \frac{1}{4}.$$

برای اینکه تحقیق کنیم فرمول (۶-۲) در مثال قبل جوابی «درست» می‌دهد، تنها کافی است در فضای نمونه‌ای کوچک‌شده‌ی A ، به دو عدد زوج احتمال U و به عدد فرد احتمال $2U$ را نسبت دهیم، به قسمی که مجموع این سه احتمال مساوی ۱ باشد.

در این صورت داریم $U + 2U + U = 1$ یا $U = \frac{1}{4}$ ، و لذا مثل قبل $P(B|A) = \frac{1}{4}$.

با ضرب طرفین فرمول تعریف ۶-۲ در $P(A)$ ، قاعده‌ی ضرب زیر را به دست می‌آوریم.

قضیه‌ی ۶-۹. اگر A و B دو پیشامد دلخواه در فضای نمونه‌ای S باشند و $P(A) \neq 0$ ، آنگاه:

$$P(A \cap B) = P(A) P(B|A).$$

با استفاده از کلمات، احتمال اینکه A و B با هم رخ دهند برابر حاصل ضرب احتمال A در احتمال شرطی B به شرط A است. در صورتی دیگر، اگر $P(B) \neq 0$ ، احتمال آنکه A و B ، هر دو رخ دهند، برابر با حاصل ضرب احتمال B و احتمال شرطی A به شرط B است. به صورت نمادی

$$P(A \cap B) = P(B) P(A|B).$$

برای به دست آوردن این قاعده‌ی حاصل ضرب، در فرمول (۶-۹) جای A و B را عوض و از واقعیت $A \cap B = B \cap A$ استفاده می‌کنیم.

مثال ۶-۲۹. اگر از محموله‌ی ۲۴۰ صفحه‌ی کلید کامپیوتر، که ۱۵ تای آن معیوب است، به تصادف و به توالی دوتای آنها را برداریم، چقدر احتمال دارد که هردو معیوب باشند؟
حل. اگر احتمال انتخاب‌ها را برابر فرض کنیم احتمال اینکه اولین صفحه‌کلید معیوب باشد $\frac{15}{240}$ ، و احتمال اینکه دومین صفحه‌کلید، به شرط معیوب بودن اولی، معیوب باشد برابر $\frac{14}{239}$ است. لذا احتمال اینکه هردو معیوب باشند $\frac{15}{240} \times \left(\frac{14}{239}\right) = \frac{7}{1912}$ است. البته فرض بر این است که انتخاب‌ها بدون جایگذاری است، یعنی اینکه قبل از انتخاب صفحه‌کلید دوم، صفحه‌کلید اول به جای خود گذاشته نمی‌شود.

مثال ۶-۳۰. پیدا کنید احتمال اینکه به تصادف، دو مهره با شماره‌ی یک به توالی از ظرفی حاوی ۴۸ مهره (در چهار رنگ و به شماره‌های یک الی دوازده) انتخاب شوند.
الف. اگر بدون جایگذاری انتخاب کنیم،
ب. اگر با جایگذاری انتخاب کنیم.
حل. الف. اگر قبل از انتخاب مهره‌ی دوم، مهره‌ی اول را به جای خود برنگردانیم، احتمال به دست آوردن متوالی دو مهره با شماره‌ی یک برابر است با:

$$\frac{4}{48} \times \frac{3}{47} = \frac{1}{188}.$$

ب. اگر قبل از انتخاب مهره‌ی دوم، مهره‌ی اول را به جای خود برگردانیم، احتمال متناظر برابر است با:

$$\frac{4}{48} \times \frac{4}{48} = \frac{1}{144}.$$

در وضعیت‌هایی که در دو مثال قبل توصیف شدند، بین دو پیشامد A و B ، یک ترتیب زمانی قطعی وجود دارد. به طور کلی، وقتی می‌نویسیم $P(A|B)$ یا $P(B|A)$ نیاز به چنین ترتیبی نیست. برای مثال، می‌توانستیم احتمال آن را بخواهیم که اولین مهره انتخاب‌شده با شماره یک باشد به شرط آنکه دومین مهره انتخاب‌شده (بدون جایگذاری) نیز شماره یک باشد- پاسخ به این سؤال نیز $\frac{3}{47}$ خواهد داد.

قضیه‌ی ۶-۹ را می‌توان به سهولت تعمیم داد، به قسمی که بتوان آن را برای بیش از دو پیشامد به کار برد. مثلاً برای سه پیشامد داریم:

قضیه‌ی ۶-۱۰. اگر A ، B ، و C سه پیشامد دلخواه در فضای نمونه‌ای S باشند، به قسمی که $P(A) \neq 0$ و $P(A \cap B) \neq 0$ ، آنگاه

$$P(A \cap B \cap C) = P(A) P(B|A) P(C|A \cap B).$$

برهان. با نوشتن $A \cap B \cap C$ به صورت $(A \cap B) \cap C$ و با دوبار استفاده از فرمول (۶-۹)، به دست می‌آوریم

$$\begin{aligned} P(A \cap B \cap C) &= P((A \cap B) \cap C) \\ &= P(A \cap B) P(C | A \cap B) \\ &= P(A) P(B|A) P(C | A \cap B). \end{aligned}$$

مثال ۶-۳۱. بسته‌ای حاوی ۲۰ فیوز است که ۵ تای آنها معیوب‌اند. اگر به تصادف ۳ فیوز به طور متوالی و بدون جایگذاری از بسته‌ی مزبور انتخاب کنیم، احتمال اینکه هر ۳ فیوز معیوب باشند چقدر است؟

حل. اگر A پیشامد معیوب بودن فیوز اول، B پیشامد معیوب بودن فیوز دوم، و C پیشامد معیوب بودن فیوز سوم باشد، آنگاه $P(A) = \frac{5}{20}$ ، $P(B|A) = \frac{4}{19}$ ، $P(C|A \cap B) = \frac{3}{18}$ ، و اگر آن را در فرمول قرار دهیم نتیجه می‌شود:

$$P(A \cap B \cap C) = \frac{5}{20} \times \frac{4}{19} \times \frac{3}{18} = \frac{1}{114}.$$

تعمیم بیشتر قضیه‌های ۶-۹ و ۶-۱۰ به k پیشامد نیز امکان‌پذیر است.

۶-۷ پیشامدهای مستقل

دو پیشامد A و B را مستقل می‌گویند در صورتی که رخ دادن یا رخ ندادن هریک از آنها در احتمال رخ دادن دیگری تأثیری نداشته باشد. برای مثال در انتخاب‌های با جایگذاری در هر مرحله نتایج از هم مستقل هستند. به‌طور نمادی، دو پیشامد A و B مستقل‌اند اگر:

$$P(A|B) = P(A), \quad P(B|A) = P(B),$$

و موقعی که هر دو احتمال شرطی موجود باشند، یعنی وقتی $P(A)$ و $P(B)$ صفر نباشند، می‌توان نشان داد که هر یک از این برابری‌ها، دیگری را نتیجه می‌دهد. حال اگر در فرمول قضیه ۶-۹، $P(B)$ را به جای $P(B|A)$ قرار دهیم، به دست می‌آوریم:

$$P(A \cap B) = P(A) P(B|A) = P(A) P(B),$$

و این نتیجه را به‌عنوان تعریف صوری استقلال به کار خواهیم برد.

تعریف ۶-۳. دو پیشامد B و C مستقل‌اند اگر و تنها اگر:

$$P(A \cap B) = P(A) P(B).$$

اگر دو پیشامد مستقل نباشند، می‌گویند وابسته‌اند. در به دست آوردن فرمول تعریف ۶-۳، فرض کردیم که $P(B|A)$ تعریف شده است و لذا فرض کردیم که $P(A) \neq 0$.

مثال ۶-۳۲. سکه‌ای همگن را سه بار پرتاب می‌کنیم فضای نمونه‌ای عبارت است از:

$$S = \{TTT, TTH, THT, HTT, THH, HTH, HHT, HHH\}.$$

فرض می‌کنیم که کلیه‌ی عناصر S هم‌شانس هستند. اگر A پیشامد رخ دادن شیر در هر دو پرتاب اول، B پیشامد رخ دادن خط در پرتاب سوم، و C پیشامد رخ دادن دقیقاً دو خط در سه پرتاب باشد، نشان دهید که:

الف. پیشامدهای A و B مستقل‌اند،

ب. پیشامدهای B و C وابسته‌اند.

حل.

$$A = \{HHH, HHT\},$$

$$B = \{HHT, HTT, THT, TTT\},$$

$$C = \{HTT, THT, TTH\},$$

$$A \cap B = \{HHT\},$$

$$B \cap C = \{HTT, THT\}.$$

فرض هم‌شانس بودن همه‌ی هشت عنصر نتیجه می‌دهد که $P(A) = \frac{1}{4}$ ، $P(B) = \frac{1}{4}$ ،

$$P(C) = \frac{3}{8}$$

$$P(A \cap B) = \frac{1}{8} \text{ و } P(B \cap C) = \frac{1}{4}$$

الف. $P(A \cap B) = \frac{1}{8}$ ، پس پیشامدهای A و B مستقل‌اند.

ب. $P(B \cap C) \neq \frac{3}{16} = \frac{1}{4} \times \frac{3}{8} = P(B)P(C)$ ، پس پیشامدهای B و C مستقل نیستند.

می‌توان نشان داد که اگر A و B مستقل باشند، آنگاه A و B' ، A' و B ، و A' و B' هم مستقل‌اند. یکی از حالت‌های ذکرشده را اثبات می‌کنیم.

قضیه‌ی ۶-۱۱. اگر دو پیشامد A و B مستقل باشند، آنگاه A و B' نیز مستقل‌اند. **برهان.** از آنجا که $A = (A \cap B) \cup (A \cap B')$ ، $A \cap B$ و $A \cap B'$ ناسازگارند و چون A و B بنابر فرض مستقل‌اند، داریم

$$\begin{aligned} P(A) &= P((A \cap B) \cup (A \cap B')) \\ &= P(A \cap B) + P(A \cap B') \\ &= P(A)P(B) + P(A \cap B'). \end{aligned}$$

نتیجه می‌شود که:

$$\begin{aligned} P(A \cap B') &= P(A) - P(A)P(B) \\ &= P(A)(1 - P(B)) \\ &= P(A)P(B'), \end{aligned}$$

و لذا A و B' مستقل‌اند.

در تمرین‌ها خواسته شده است که نشان دهید اگر A و B مستقل باشند، آنگاه A' و B ، و A' و B' نیز مستقل‌اند، و اگر A و B وابسته باشند، آنگاه A و B' وابسته‌اند.

برای تعمیم مفهوم استقلال به بیش از دو پیشامد، تعریف زیر را ارائه می‌دهیم.

تعریف ۶-۴: پیشامدهای $A_1, A_2, \dots, A_k$ مستقل اند اگر و تنها اگر احتمال اشتراک هر ۲، ۳، ...، k تا از این پیشامدها مساوی حاصل ضرب احتمال‌های مربوط به هر پیشامد باشد.

برای حالتی که سه پیشامد A ، B و C مطرح است شرایط استقلال به صورت زیر خواهد بود:

$$\begin{aligned} P(A \cap B) &= P(A) P(B), \\ P(A \cap C) &= P(A) P(C), \\ P(B \cap C) &= P(B) P(C), \end{aligned}$$

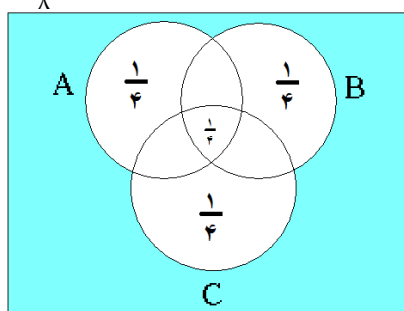
و

$$P(A \cap B \cap C) = P(A) P(B) P(C).$$

قابل توجه است که سه یا چند پیشامد می‌توانند دوبه‌دو مستقل باشند بدون آنکه کلاً مستقل باشند.

مثال ۶-۳۳: سه پیشامد با احتمال‌های مشخص شده در نمودار ون شکل ۶-۲، آمده است. تحقیق کنید که A و B مستقل اند، A و C مستقل اند، B و C مستقل اند، اما A ، B و C مستقل نیستند.

حل. همان‌طور که از نمودار دیده می‌شود، $P(A) = P(B) = P(C) = \frac{1}{4}$ ، پس $P(A \cap B) = P(A \cap C) = P(B \cap C) = \frac{1}{4}$ و $P(A \cap B \cap C) = \frac{1}{8}$.
 $P(A) P(B) = \frac{1}{4} = P(A \cap B)$ ، $P(A) P(C) = \frac{1}{4} = P(A \cap C)$ ،
 $P(B) P(C) = \frac{1}{4} = P(B \cap C)$ ، اما $P(A) P(B) P(C) = \frac{1}{8} \neq P(A \cap B \cap C)$.



شکل ۶-۲. نمودار ون برای مثال ۶-۳۳

ممکن است که $P(A \cap B \cap C) = P(A)P(B)P(C)$ بدون آنکه A ، B ، و C دوه‌دو مستقل باشند. مثال زیر را ببینید.

مثال ۶-۳۴. مطلوب است احتمال به‌دست آوردن:

الف. سه شیر در سه پرتاب تصادفی یک سکه‌ی همگن،

ب. چهار شش و سپس عددی دیگر در پنج بار ریختن یک تاس همگن.

حل. الف. از ضرب احتمال‌های مربوط، به‌دست می‌آوریم:

$$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}.$$

ب. از ضرب احتمال‌های مربوط، به‌دست می‌آوریم:

$$\frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} \times \frac{5}{6} = \frac{5}{7776}.$$

۶-۸ قضیه‌ی بیز

مسائل زیادی وجود دارند که در آنها برآمد نهایی یک آزمایش به آنچه در مراحل میانی مختلف رخ می‌دهند بستگی دارد. مثال زیر، مثالی ساده است که در آن، مرحله‌ای میانی وجود دارد که متشکل از دو شق است.

مثال ۶-۳۵. کار تکمیل بنای بزرگراهی ممکن است به علت اعتصاب به تأخیر افتد. فرض کنید احتمال این که اعتصابی رخ دهد 0.60 ، احتمال این که اگر اعتصابی نباشد کار به موقع انجام شود 0.85 ، و احتمال این که اگر اعتصابی باشد کار به موقع انجام شود، 0.35 باشد. احتمال این که کار به موقع انجام شود چقدر است؟

حل. اگر A پیشامد تکمیل به‌موقع کار، و B پیشامد وقوع اعتصاب باشد، اطلاعات داده شده را می‌توان چنین نوشت:

$$P(B) = 0.60, P(A|B) = 0.35, P(A|B') = 0.85.$$

با استفاده از قضیه‌های احتمال، و این واقعیت که $A \cap B$ و $A \cap B'$ ناسازگارند، و صورت دیگر قاعده‌ی ضرب، می‌توانیم بنویسیم:

$$\begin{aligned}
 P(A) &= P((A \cap B) \cup (A \cap B')) \\
 &= P(A \cap B) + P(A \cap B') \\
 &= P(B) P(A|B) + P(B') P(A|B').
 \end{aligned}$$

در این صورت، با قراردادن مقادیر عددی، نتیجه می‌شود که:

$$P(A) = (0,60)(0,35) + (1-0,60)(0,85) = 0,55.$$

تعمیمی از این نوع وضعیت، موردی است که در آن مرحله‌ی میانی دارای k شق مختلف است (که رخ دادن آنها را با $B_1, B_2, \dots, B_k$ نشان می‌دهیم). این تعمیم، به قضیه‌ی ۶-۱۲ که گاهی آن را **قاعده‌ی احتمال کل** یا **قاعده‌ی حذف** می‌نامند نیاز دارد.

قضیه‌ی ۶-۱۲. اگر پیشامدهای $B_1, B_2, \dots, B_k$ افزازی از فضای نمونه‌ای S را تشکیل دهند و به ازای هر i ، $i=1, 2, \dots, k$ ، $P(B_i) \neq 0$ ، آنگاه برای هر پیشامد A در S

$$P(A) = \sum_{i=1}^k P(B_i) P(A|B_i).$$

در اینجا B ها افزازی از فضای نمونه‌ای را تشکیل می‌دهند، اگر دوبه‌دو ناسازگار بوده و اجتماع آنها مساوی S باشد.

مثال ۶-۳۶. اعضای یک شرکت مشاور، از سه آژانس، اتومبیل کرایه می‌کنند: از آژانس ۱ به میزان ۶۰ درصد، از آژانس ۲ به میزان ۳۰ درصد، و از آژانس ۳ به میزان ۱۰ درصد. اگر ۹ درصد از اتومبیل‌های آژانس ۱، ۲۰ درصد از اتومبیل‌های آژانس ۲، و ۶ درصد از اتومبیل‌های آژانس ۳ به تنظیم موتور نیاز داشته باشند، احتمال اینکه یک اتومبیل کرایه‌ای که به شرکت مشاور تحویل شده است به تنظیم موتور نیاز داشته باشد چقدر است؟

حل. اگر A پیشامد نیاز داشتن اتومبیلی به تنظیم موتور و B_1, B_2, B_3 به ترتیب پیشامدهای تعلق یک اتومبیل کرایه‌شده به آژانس‌های ۱، ۲، ۳ باشند، داریم:

$$\begin{aligned}
 P(B_1) &= 0,60, P(B_2) = 0,30, P(B_3) = 0,10, \\
 P(A|B_1) &= 0,09, P(A|B_2) = 0,20, P(A|B_3) = 0,06.
 \end{aligned}$$

اگر این مقادیر را در فرمول قضیه‌ی ۶-۱۲ قرار دهیم، نتیجه می‌شود که:

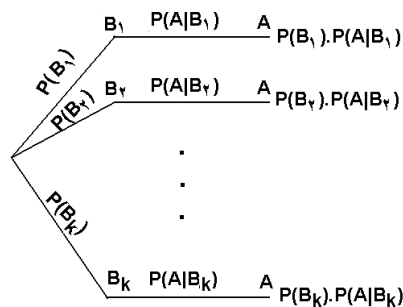
$$P(A) = (0,60)(0,09) + (0,30)(0,20) + (0,10)(0,6) = 0,12.$$

پس، ۱۲ درصد از اتومبیل‌هایی که به شرکت تحویل می‌شوند به تنظیم موتور نیاز دارند. با رجوع به مثال قبلی، فرض کنیم سؤال زیر موردنظر باشد: اگر یک اتومبیل کرایه که به شرکت مشاور تحویل شده است نیاز به تنظیم موتور داشته باشد، چقدر احتمال دارد که متعلق به آژانس ۲ باشد؟ برای پاسخ‌دادن به این نوع سؤال‌ها، به قضیه‌ی زیر که **قضیه‌ی بیز** نامیده می‌شود نیاز داریم:

قضیه‌ی ۶-۱۳. اگر پیشامدهای $B_1, B_2, \dots, B_k$ افزای از فضای نمونه‌ای S را تشکیل دهند و به ازای هر i ، $i = 1, 2, \dots, k$ ، $P(B_i) \neq 0$ ، آنگاه برای هر پیشامد A در S ، به قسمی که $P(A) \neq 0$ ،

$$P(B_r | A) = \frac{P(B_r) P(A | B_r)}{\sum_{i=1}^k P(B_i) P(A | B_i)} \quad r = 1, 2, \dots, k.$$

با استفاده از کلمات، احتمال آنکه از راه شاخه‌ی r نمودار درختی شکل ۶-۳ به پیشامد A برسیم، به شرط آنکه از راه یکی از k شاخه‌ی نمودار به A رسیده باشیم، برابر نسبت احتمال مربوط به r مین شاخه به مجموع احتمال‌های مربوط به کل k شاخه‌ی درخت است.



شکل ۶-۳. نمودار درختی برای قضیه‌ی بیز

برهان. مطابق تعریف احتمال شرطی، می‌نویسیم $P(B_r | A) = \frac{P(A \cap B_r)}{P(A)}$. لازم است فقط به جای $P(A \cap B_r)$ مقدار $P(B_r)P(A | B_r)$ ، و به جای $P(A)$ عبارت فرمول قضیه‌ی ۶-۱۲ را قرار دهیم.

مثال ۶-۳۷. با مراجعه به مثال ۶-۳۶، اگر اتومبیل کرایه‌ای که به شرکت مشاور تحویل شده است به تنظیم موتور نیاز داشته باشد، احتمال اینکه متعلق به آژانس ۲ باشد چقدر است؟

حل. با توجه به فرمول قضیه‌ی (۶-۱۳)، قرار می‌دهیم و داریم:

$$P(B_2 | A) = \frac{(0.30)(0.20)}{(0.60)(0.90) + (0.30)(0.20) + (0.10)(0.6)} \\ = \frac{0.06}{0.12} = 0.5.$$

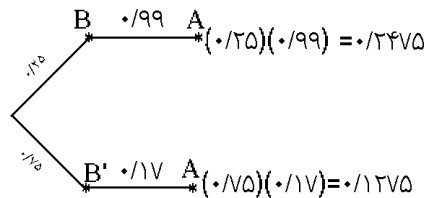
توجه کنید که گرچه فقط ۳۰ درصد اتومبیل‌هایی که به شرکت کرایه داده می‌شوند متعلق به آژانس ۲ است، ولی ۵۰ درصد اتومبیل‌هایی که به تنظیم نیاز دارند متعلق به این آژانس‌اند.

مثال ۶-۳۸. در استان معینی، دودزایی ۲۵ درصد تمام اتومبیل‌ها بیش از حد عادی است. اگر در آزمون دودزایی وسایل نقلیه‌ی استان، اتومبیلی که بیش از حد دود تولید می‌کند با احتمالی برای ۰/۹۹ مردود شود و اگر اتومبیلی که دودزایی آن بیش از حد عادی نیست با احتمال ۰/۱۷ در این آزمون رد شود، احتمال اینکه اتومبیلی که در آزمون رد می‌شود واقعاً اتومبیلی باشد که دودزایی آن بیش از حد عادی است چقدر است؟

حل. از شکل ۶-۴ درمی‌یابیم که احتمال‌های مربوط به دو شاخه‌ی نمودار درختی عبارت‌اند از $0.2475 = (0.99)(0.25)$ و $0.1275 = (0.17)(1 - 0.25)$ ، پس احتمال اینکه اتومبیلی که در آزمون رد می‌شود، واقعاً اتومبیلی باشد که بیش از حد دود ایجاد می‌کند برابر است با:

$$\frac{0.2475}{0.2475 + 0.1275} = 0.66.$$

البته می‌توانستیم این نتیجه را از قراردادن مستقیم در فرمول قضیه‌ی بیز، بدون نمودار درختی نیز به دست آوریم.



شکل ۶-۴. نمودار درختی برای مثال ۶-۳۸

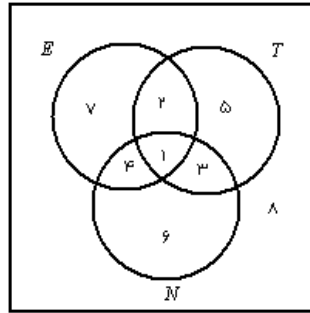
گرچه قضیه‌ی بیز از اصول احتمال و تعریف احتمال شرطی نتیجه می‌شود، ولی این قضیه موضوع مباحثه‌های دامنه‌داری بوده است. درباره‌ی اعتبار قضیه‌ی بیز نمی‌توان سؤالی مطرح کرد، اما بحث‌های مفصلی درباره‌ی تعبیر احتمال‌های پیشین $P(B_i)$ مطرح شده‌اند. همچنین درصدی از هاله‌ی اسراری که قضیه‌ی بیز را احاطه کرده‌است، ناشی از این واقعیت است که این قضیه نظیر مثال ۶-۳۸، نوعی استدلال «قهقراپی» یا «معکوس»، یعنی استدلال «از معلول به علت» را موجب می‌شود. مثلاً در مثال ۶-۳۸، رد شدن از آزمون، یک معلول، و دودزایی بیش از حد، یک علت ممکن است.

خلاصه‌ی فصل ششم

در ابتدای فصل، شرایط آزمایش تصادفی بیان شد و مجموعه‌ی کلیه‌ی برآمدهای ممکنه از یک آزمایش تصادفی، فضای نمونه‌ای تعریف شد. زیرمجموعه‌ی فضای نمونه‌ای پیشامد نام گرفت. با توجه به نوع فضای نمونه‌ای، که ممکن است شمارش‌پذیر متناهی و یا نامتناهی و یا غیرقابل شمارش‌پذیر باشد، حالت‌های مختلفی نیز برای پیشامدها برقرار است. عملیات مختلف روی پیشامدها، مانند اجتماع، اشتراک، متمم‌گیری و ... قابل انجام است. مفهوم سازگاری برای دو پیشامد که اشتراکشان تهی است تعریف می‌شود. احتمال شرطی، که در آن با دانستن وقوع یک پیشامد، احتمال وقوع دیگری محاسبه می‌شود، مطرح شد. حالتی که اطلاعات قسمت شرط سودمند نیست وقتی است که پیشامدها مستقل باشند؛ یعنی حاصل ضرب احتمال تک‌تک آنها با احتمال اشتراک آنها برابر است. برخی آزمایش‌ها طی دو مرحله یا بیش از دو مرحله انجام می‌شوند که در این حالت، قضیه‌ی بیز و استفاده از نمودار درختی پیشنهاد می‌شود.

تمرین

- ۶-۱. فضای نمونه‌ای حاصل از پرتاب دو سکه را بنویسید.
- ۶-۲. بر روی قوطی کنسرو لوییا، عبارت «وزن 250 ± 5 گرم» چاپ شده است، فضای نمونه‌ای وزن قوطی‌های کنسرو را بنویسید.
- ۶-۳. مقاومت‌های برشی دو نقطه‌ای جوش باید اندازه‌گیری شود. اگر حد بالایی را با U مشخص کنیم، فضای نمونه‌ای مقاومت برشی دو نقطه‌ای جوش را بنویسید.
- ۶-۴. برای تعداد بازدیدکنندگان هر گروه از اثر باستانی پاسارگاد در ایام نوروز، فضای نمونه‌ای مناسبی بنویسید.
- ۶-۵. در ساخت بال یک هواپیما از تعداد زیادی میخ پرچ استفاده می‌شود. فضای نمونه‌ای تعداد میخ پرچ‌های خراب در بازرسی یک بال هواپیما را بنویسید.
- ۶-۶. سکه‌ای را یک بار پرتاب می‌کنیم، اگر شیر بیاید، تاسی را یک بار می‌ریزیم، اگر خط بیاید، سکه را دوبار دیگر پرتاب می‌کنیم. فضای نمونه‌ای مربوط را بنویسید.
- ۶-۷. اگر $S = \{x | 0 < x < 10\}$ ، $A = \{x | 3 < x \leq 8\}$ ، و $B = \{x | 5 < x < 10\}$ ، پیشامدهای زیر را تعیین کنید.
الف. $A \cup B$ ب. $A \cap B$ ج. $A \cap B'$ د. $A' \cup B$
- ۶-۸. از گروه ۲۰۰ نفری دانشجویان یک دانشکده، ۱۳۸ نفر درس آمار، ۱۱۵ نفر درس ریاضی عمومی، و ۹۱ فرد در هر دو درس ثبت‌نام کرده‌اند. چند نفر از این دانشجویان در هیچ یک از این دروس ثبت‌نام نکرده‌اند؟
- ۶-۹. در شکل ۶-۵، E ، T ، و N پیشامدهایی به ترتیب هستند که خودروهایی که به تعمیرگاهی آورده‌اند نیاز به تعمیر کامل موتور، تعمیر جعبه‌دنده، یا تایرهای نو دارند. پیشامدهایی را که با
- الف. ناحیه ۱، ب. ناحیه ۳، ج. ناحیه ۷، د. نواحی ۱ و ۴ باهم، ه. نواحی ۲ و ۵ باهم، و. نواحی ۳، ۵، و ۶ باهم، نشان داده می‌شوند، در قالب کلمات بیان کنید.



شکل ۶-۵. شکل مورد نیاز برای تمرین ۶-۹

۶-۱۰. با رجوع به تمرین ۶-۹ و شکل مربوطه، ناحیه یا ترکیبی از نواحی را فهرست کنید که معرف پیشامدهایی برای خودروهایی توصیف شده باشد.

الف. تعمیر جعبه‌دنده، ولی نه تعمیر کامل موتور و نه نیاز به تایر نو،

ب. تعمیر کامل موتور و تعمیر جعبه‌دنده،

ج. تعمیر جعبه‌دنده یا نیاز به تایر نو ولی نه تعمیر کامل موتور،

د. نیاز به تایر نو.

۶-۱۱. اگر A و B دو پیشامد در فضای نمونه‌ای S باشند و $P(A) = 0.2$ ، $P(B) = 0.6$ و $P(A \cup B) = 0.7$ ، مقادیر $P(A \cap B)$ ، $P(A' \cap B)$ ، $P(A' \cup B')$ و $P(A' \cap B')$ را به دست آورید.

۶-۱۲. فرمولی برای محاسبه $P(A \Delta B)$ برحسب $P(A)$ ، $P(B)$ و $P(A \cap B)$ ارائه دهید (احتمال اینکه دقیقاً یکی از دو پیشامد رخ دهد).

۶-۱۳. اگر A و B دو پیشامد در فضای نمونه‌ای S باشند، احتمال پیشامدهای زیر را برحسب $P(A)$ ، $P(B)$ و $P(A \cap B)$ به دست آورید:

الف. دقیقاً i تا، $i = 0, 1, 2$ ، از پیشامدهای A و B رخ دهند.

ب. حداقل i تا، $i = 0, 1, 2$ ، از پیشامدهای A و B رخ دهند.

ج. حداکثر i تا، $i = 0, 1, 2$ ، از پیشامدهای A و B رخ دهند.

۶-۱۴. جعبه‌ای دارای ۲۴ مهره است که از ۱ تا ۲۴ شماره‌گذاری شده است. یک مهره به تصادف از این جعبه انتخاب می‌کنیم. اگر A پیشامد مشاهده‌ی عددی بخش‌پذیر بر

۲ و B پیشامد مشاهده‌ی عددی بخش‌پذیر بر ۳ باشد، مقدار $P(A \cap B)$ را به دست آورید.

۱۵-۶. برای A و B دو پیشامد در فضای نمونه‌ای S ، نشان دهید

$$\text{الف. } P(A \cap B) \leq P(A) + P(B)$$

$$\text{ب. } P(A \cap B) \geq P(A) + P(B) - 1$$

۱۶-۶. به کمک استقرا ثابت کنید که برای هر دنباله‌ی متناهی از پیشامدهای

$$E_1, E_2, \dots, E_n$$

$$P(E_1 \cup E_2 \cup \dots \cup E_n) \leq \sum_{i=1}^n P(E_i).$$

۱۷-۶. بخت رخ دادن یک پیشامد برحسب نسبت احتمال وقوع آن به احتمال عدم وقوع آن داده می‌شود، به شرط آنکه هیچ یک از دو احتمال صفر نباشد. بخت را معمولاً برحسب دو عدد صحیح مثبت که عامل مشترک ندارند، بیان می‌کنند. اگر بخت اینکه پیشامدی رخ دهد برابر a به b باشد، نشان دهید که احتمال این پیشامد $p = \frac{a}{a+b}$ است.

۱۸-۶. توضیح دهید که چرا در هریک از احکام زیر اشتباهی وجود دارد:

الف. احتمال اینکه فرهاد در امتحان وکالت قبول شود $0/66$ و احتمال اینکه رد شود $0/34$ است.

ب. احتمال اینکه یک منشی در تایپ گزارشی مرتکب ۰، ۱، ۲، ۳، و ۴ یا بیش از ۵ اشتباه شود به ترتیب $0/12$ ، $0/25$ ، $0/36$ ، $0/14$ ، $0/09$ ، $0/07$ است.

ج. احتمال این که بانکی در روزی مفروض ۰، ۱، ۲ یا بیش از ۳ چک بدون محل دریافت کند به ترتیب $0/08$ ، $0/21$ ، $0/29$ ، $0/40$ است.

۱۹-۶. احتمال اینکه نحوه‌ی بهره‌برداری از یک دستگاه جدید رادیولوژی، به «خیلی سخت»، «سخت»، «متوسط»، «آسان» و «خیلی آسان» درجه‌بندی شود به ترتیب $0/17$ ، $0/34$ ، $0/29$ ، $0/08$ است. پیدا کنید احتمال اینکه نحوه‌ی بهره‌برداری دستگاه به یکی از صورتهای زیر درجه‌بندی شود:

الف. سخت یا خیلی سخت، ب. نه خیلی سخت و نه خیلی آسان،

ج. متوسط یا بدتر، د. متوسط یا بهتر.

۶-۲۰. شرکت ایران خودرو برای تأمین لاستیک، می‌تواند با احتمال‌های 0.15 ، 0.24 ، 0.03 ، 0.28 ، 0.22 ، 0.08 به ترتیب از شرکت‌های A ، B ، C ، D ، E ، یا F خریداری کند. احتمال خرید لاستیک را در هریک از حالت‌های زیر تعیین کنید.

الف. از شرکت B یا E باشد، ب. از شرکت A ، C یا E باشد،

ج. از شرکت C یا F باشد، د. از شرکت A ، C ، D یا E باشد.

۶-۲۱. در کلاهی بیست تکه کاغذ سفید که از ۱ تا ۲۰، ده تکه کاغذ قرمز که از ۱ تا ۱۰، چهل تکه کاغذ زرد که از ۱ تا ۴۰، و ده تکه کاغذ آبی که از ۱ تا ۱۰ شماره‌گذاری شده‌اند ریخته شده است. اگر این ۸۰ تکه کاغذ را کاملاً مخلوط کنیم به‌قسمی که تمام تکه کاغذها در موقع استخراج احتمال برابر داشته باشند، پیدا کنید احتمال استخراج یک تکه کاغذ را که:

الف. آبی یا سفید باشد،

ب. به‌شماره‌ی ۱، ۲، ۳، ۴، یا ۵ باشد،

ج. قرمز یا زرد و با شماره‌ی ۱، ۲، ۳، یا ۴ باشد،

د. به‌شماره‌ی ۵، ۱۵، ۲۵، ۳۵ باشد،

ه. سفید و با شماره‌ای بزرگتر از ۱۲ یا زرد و با شماره‌ی بزرگتر از ۲۶ باشد.

۶-۲۲. یک استاد زیست‌شناسی در کارهای تحقیقاتی‌اش دو دستیار دارد. احتمال اینکه دستیار مسن‌تر در روز معینی غیبت کند 0.08 ، و احتمال اینکه دستیار جوان‌تر در روز معینی غیبت کند 0.05 ، و احتمال اینکه هر دو در روز معینی غایب باشند 0.02 است. پیدا کنید احتمال اینکه:

الف. حداکثر یکی از هر دو دستیار در روز معینی غایب باشند،

ب. حداقل یکی از دو دستیار در روز معینی غایب نباشد،

ج. فقط یکی از دو دستیار در روز معینی غایب باشد.

۶-۲۳. نشان دهید که احتمال‌های شرطی در سه اصل موضوع احتمال صادقند، به عبارت دیگر نشان دهید که اگر $P(B) \neq 0$ آنگاه

الف. $P(A|B) \geq 0$ ،

ب. $P(B|B) = 1$ ،

ج. برای هر دنباله از پیشامدهای دوبه‌دو ناسازگار $A_1, A_2, \dots$

$$P(A \cup A_1 \cup \dots | B) = P(A | B) + P(A_1 | B) + \dots$$

۶-۲۴. به کمک مثال‌های عددی نشان دهید که $P(B|A) + P(B|A') \neq 1$

الف. ممکن است مساوی ۱ باشد،

ب. لازم نیست مساوی ۱ باشد.

۶-۲۵. نشان دهید که به شرط $P(A \cap B \cap C) \neq 0$

$$P(A \cap B \cap C \cap D) = P(A) \cdot P(B|A) \cdot P(C|A \cap B) \cdot P(D|A \cap B \cap C).$$

۶-۲۶. برای سه پیشامد A ، B ، و C ، با داشتن شرایط

$$P(C|A \cap B) = P(C|B) \text{ و } P(A \cap B \cap C) \neq 0$$

نشان دهید که $P(A|B \cap C) = P(A|B)$.

۶-۲۷. برای تصدی شغلی در بخش خبری یک ایستگاه تلویزیون، ۹۰ داوطلب وجود

دارند. بعضی از آنها فارغ‌التحصیل دانشگاه‌اند، و برخی نیستند، برخی از آنها حداقل سه

سال سابقه کار دارند و برخی ندارند. جدولی تفکیکی به صورت جدول ۶-۲ است:

جدول ۶-۲. اطلاعات تمرین ۶-۲۷

	فارغ‌التحصیل دانشگاه	دانشگاه نرفته
حداقل با سه سال تجربه	۱۸	۹
کمتر از سه سال تجربه	۳۶	۲۷

اگر ترتیب مصاحبه‌ی داوطلبان به وسیله‌ی مدیر ایستگاه، تصادفی باشد، G پیشامدی

است که اولین داوطلب که مصاحبه می‌شود فارغ‌التحصیل دانشگاه است، و T

پیشامدی است که اولین داوطلبی که مصاحبه می‌شود حداقل سه سال سابقه‌ی کار دارد.

هر یک از احتمال‌های زیر را مستقیماً از درایه‌ها و حاصل جمع‌های سطری و ستونی

جدول تعیین کنید.

الف. $P(G)$ ، ب. $P(T')$ ، ج. $P(G \cap T)$

د. $P(G' \cap T')$ ، ه. $P(T|G)$ ، و. $P(G'|T')$

۶-۲۸. نتایج تمرین قبل را برای تحقیق درستی تساوی‌های زیر به کار برید.

$$\text{الف. } P(T|G) = \frac{P(G \cap T)}{P(G)} \quad \text{ب. } P(G'|T') = \frac{P(G' \cap T')}{P(T')}$$

۶-۲۹. نشان دهید که اگر $P(B|A) = P(B)$ و $P(B) \neq 0$ ، آنگاه

$$P(A|B) = P(A)$$

۶-۳۰. جعبه‌ای شامل ۴ مهره‌ی سفید و ۸ مهره‌ی سیاه است. ۵ مهره به تصادف و بدون جایگذاری از این جعبه و بدون نگاه کردن به رنگ مهره‌ها کنار گذاشته می‌شود.

الف. اگر یک مهره مجدداً از این جعبه انتخاب شود، احتمال اینکه مهره‌ی انتخابی سفید باشد چقدر است؟

ب. اگر دو مهره مجدداً به تصادف از این جعبه انتخاب شود، احتمال اینکه دقیقاً شامل یک مهره‌ی سفید باشد چقدر است؟

۶-۳۱. جعبه‌ای شامل ۸ مهره‌ی سفید و ۴ مهره‌ی سیاه است. یک نمونه‌ی تصادفی ۴تایی بدون جایگذاری از این جعبه انتخاب می‌شود.

الف. احتمال این را که مهره‌های دوم و سوم سفید باشند در حالی که بدانیم نمونه دقیقاً شامل ۳ مهره‌ی سفید است به دست آورید.

ب. احتمال این را که مهره‌ی اول سفید باشد در حالی که بدانیم نمونه شامل ۲ مهره‌ی سفید است به دست آورید.

۶-۳۲. سه جعبه را در نظر بگیرید، به طوری که در جعبه‌ی اول ۳ مهره‌ی سفید و ۵ مهره‌ی سیاه، در جعبه‌ی دوم ۶ مهره‌ی سفید و ۴ مهره‌ی سیاه و در جعبه‌ی سوم ۲ مهره‌ی سفید و ۴ مهره‌ی سیاه است. اگر یک مهره به تصادف از هر جعبه انتخاب شود، احتمال اینکه مهره‌ی انتخابی از جعبه‌ی اول سفید باشد به شرط آن که دقیقاً ۲ مهره‌ی سفید انتخاب شده باشد، چقدر است؟

۶-۳۳. نشان دهید که اگر دو پیشامد A و B مستقل باشند، آنگاه:

الف. دو پیشامد A' و B مستقل اند،

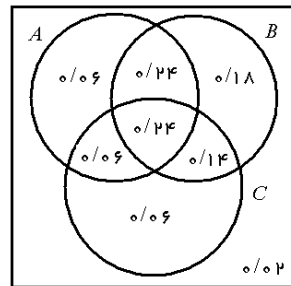
ب. دو پیشامد A' و B' مستقل اند.

۶-۳۴. نشان دهید که اگر پیشامدهای A و B وابسته باشند، آنگاه پیشامدهای A و B' وابسته اند.

۶-۳۵. با رجوع به شکل ۶-۶ نشان دهید که $P(A \cap B \cap C) = P(A)P(B)P(C)$ ، الزاماً نتیجه نمی‌دهد که پیشامدهای A ، B ، و C همگی دوه‌دو مستقلند.

۶-۳۶. با رجوع به شکل ۶-۶ نشان دهید که اگر A از B مستقل و A از C مستقل باشد، آنگاه B الزاماً از C مستقل نیست.

۶-۳۷. با رجوع به شکل ۶-۶ نشان دهید که اگر A از B مستقل و A از C مستقل باشد، آنگاه A الزاماً از $B \cup C$ مستقل نیست.



شکل ۶-۶. شکل مورد نیاز برای تمرین‌های ۶-۳۵، ۶-۳۶، و ۶-۳۷

۶-۳۸. اگر سه پیشامد A ، B ، و C مستقل باشند، نشان دهید که

الف. A و $B \cap C$ مستقل‌اند، ب. A و $B \cup C$ مستقل‌اند،

۶-۳۹. نشان دهید برای اینکه k پیشامد مستقل باشد، باید $2^k - k - 1$ شرط برقرار باشد.

۶-۴۰. سکه ای به قسمی دستکاری شده است که احتمال آمدن شیر و خط به ترتیب

0.52 و 0.48 است. اگر این سکه سه بار پرتاب شود، احتمال به دست آوردن

الف. سه شیر،

ب. دو خط و یک شیر با همین ترتیب.

۶-۴۱. اگر پنج کامیون از ده کامیون حمل بار شرکتی، استانداردهای مربوط به میزان

دود آگزوز را نداشته باشند و سه تا از آنها برای بررسی انتخاب شوند، احتمال آنکه

هیچ از یک کامیون‌های منتخب استانداردهای مذکور را نداشته باشد چقدر است؟

۶-۴۲. اگر شخصی از ۱۵ سکه‌ی طلای موجودی یک صراف که ۶ تایی آنها تقلبی

است ۴ سکه به تصادف انتخاب کند، احتمال آنکه سکه‌های منتخب همه تقلبی باشند

چقدر است؟

۶-۴۳. یک فروشگاه بزرگ، ماهیانه، صورت حساب بدهی مشتریانی را که حق نسیه بردن دارند برای آنها می فرستند. فروشگاه دریافته است مشتری ای که بلافاصله در یک ماه بدهی خود را می پردازند با احتمال $0/9$ ماه بعد نیز بلافاصله بدهی خود را می پردازند، اما اگر در یک ماه بدهی خود را بلافاصله نپردازد احتمال این که در ماه بعد بلافاصله بدهی اش را بپردازد $0/4$ است.

الف. مطلوب است احتمال آنکه مشتری ای که در ماهی بلافاصله بدهی خود را پرداخته است در سه ماه بعد نیز بلافاصله بدهی اش را بپردازد؟

ب. مطلوب است احتمال آنکه مشتری ای که در ماهی بلافاصله بدهی خود را نپرداخته است در دو ماه بعد نیز بدهی اش را فوراً نپردازد و آنگاه ماه بعد از آن فوراً بدهی خود را بپردازد.

۶-۴۴. یک کارخانه ی خودروسازی، بلبرینگ های مورد نیاز خود را از سه تولیدکننده به شرح: 50% از تولیدکننده ی ۱، 30% از تولیدکننده ی ۲ و 20% از تولیدکننده ی ۳، تهیه می کند. براساس تجربه ی گذشته شرکت خودروسازی می داند که استاندارد کنترل کیفیت این سه تولیدکننده یکسان نیست، به طوری که 2% از محصولات تولیدی ۱، 2% از محصولات تولیدی ۲، و 4% از محصولات تولیدی ۳، معیوب هستند. چه نسبتی از بلبرینگ های خریداری شده ی کارخانه خودروسازی، معیوب هستند.

۶-۴۵. در تمرین ۶-۴۴، اگر بلبرینگ انتخابی معیوب باشد، احتمال اینکه بلبرینگ از محصولات تولیدی ۲ باشد چقدر است؟

۶-۴۶. جعبه ی I شامل ۳ مهره سفید و ۵ مهره سیاه است و جعبه ی II شامل یک مهره سفید و یک مهره سیاه است. دو مهره به تصادف از جعبه ی I انتخاب و بدون نگاه کردن به جعبه ی II منتقل می کنیم و یک مهره به تصادف از جعبه ی II انتخاب می کنیم.

الف. احتمال اینکه مهره ی سفید از جعبه II استخراج شده باشد چقدر است؟
 ب. احتمال انتقال مهره های سفید از جعبه ی I به II، به شرط مشاهده ی مهره ی سفید از جعبه ی II، چقدر است؟

۶-۴۷. جعبه‌ی I شامل ۲ مهره‌ی سفید و ۴ مهره‌ی سیاه است و جعبه‌ی II شامل ۳ مهره‌ی سفید و ۲ مهره‌ی سیاه است. یک جعبه به تصادف انتخاب و دو مهره از این جعبه انتخاب می‌شود.

الف. احتمال اینکه هر دو مهره سفید باشند، چقدر است؟

ب. احتمال اینکه جعبه‌ی II انتخاب شود به شرط آنکه هر دو مهره سفید باشند چقدر است؟

خودآزمایی چهارگزینه‌ای

۶-۱. تاسی را شش بار پرتاب می‌کنیم. پیشامد A مشاهده‌ی حداقل یک بار عدد شش، و B مشاهده‌ی عدد شش در مرتبه‌های اول، دوم، و سوم است. تفاضل متقارن این دو پیشامد با کدام گزینه یکی است؟

الف. $A - B$ ب. A ج. B د. $A \cap B$

۶-۲. در جعبه‌ای که حاوی ۴ مهره قرمز و ۵ مهره سیاه است، احتمال آنکه مهره‌ی قرمز انتخاب شود، چقدر است؟

الف. $\frac{4}{9}$ ب. $\frac{1}{4}$ ج. $\frac{1}{5}$ د. $\frac{1}{9}$

۶-۳. کدام یک از گزینه‌ها جزء اصول موضوع احتمال نیستند؟

الف. $0 \leq P(A) \leq 1$, $A \subset S$

ب. $P(S) = 1$

ج. برای هر دو پیشامد ناسازگار $P(A \cup B) = P(A) + P(B)$

د. اگر $A \subset B$ آنگاه $P(A) \leq P(B)$

۶-۴. در یک بسته ۴۸ پیراهن وجود دارد (پیراهن‌ها در چهار رنگ در دوازده اندازه) پنج پیراهن به تصادف اختیار می‌کنیم. احتمال آنکه سه پیراهن دارای یک اندازه و دو پیراهن دیگر هم دارای یک اندازه باشند، چقدر است؟

الف. $\frac{12 \times 11}{\binom{48}{5}}$ ب. $\frac{4 \times \binom{4}{2}}{\binom{48}{5}}$ ج. $\frac{12 \times 11 \times \binom{4}{3} \times \binom{4}{2}}{\binom{48}{5}}$ د. $\frac{\binom{12}{3} \times \binom{11}{2}}{\binom{48}{5}}$

۶-۵. با فرض اینکه A و B دو پیشامد ناسازگار هستند و $P(A) = 0.5$ و

$P(B) = 0.2$ ، آنگاه مقدار $P(A-B)$ برابر کدام گزینه است؟

الف. ۰.۵ ب. ۰.۲ ج. ۰.۳ د. ۰.۴

۶-۶. در ظرف اول a مهره آبی، و b مهره سفید، و در ظرف دوم $a+b$ مهره

آبی، و b مهره سفید وجود دارد. از ظرف اول مهره‌ای به تصادف انتخاب نموده در

ظرف دوم قرار می‌دهیم. احتمال آنکه در بار دوم مهره آبی باشد چقدر است؟

الف. $\frac{a+b}{a+2b}$ ب. $\frac{a}{b}$ ج. $\frac{a(a+b)+b(a+b+1)}{(a+b)(a+2b+1)}$ د. $\frac{a}{a+b}$

۶-۷. در ظرف اول a مهره آبی، و b مهره سفید، و در ظرف دوم $a+b$ مهره

آبی، و b مهره سفید وجود دارد. از ظرف اول مهره‌ای به تصادف انتخاب نموده در

ظرف دوم قرار می‌دهیم. اگر بدانیم مهره انتخابی در بار دوم آبی است، شانس آنکه

در بار اول نیز مهره آبی بوده، چقدر است؟

الف. $\frac{a(a+b)+b(a+b+1)}{(a+b)(a+2b+1)}$ ب. $\frac{b(a+2b)}{a(a+b)+b(a+2b+1)}$ ج. $\frac{b}{a(a+b)+b}$ د. $\frac{b(a+b+1)}{a(a+b)+b(a+2b+1)}$

۶-۸. رابطه‌ی $P(A \cup B) = P(A) + P(B)$ در چه شرطی برقرار است؟

الف. همواره برقرار است.

ب. برای دو پیشامد مستقل برقرار است.

ج. برای دو پیشامد ناسازگار برقرار است.

د. برای دو پیشامد غیرتهی برقرار است.

۶-۹. اگر $P(A) = P(B) = P(C) = x$ و سه پیشامد مستقل باشند، آنگاه مقدار

$P(A \cup B \cup C)$ برابر کدام گزینه است؟

الف. $3x - 3x^2 + x^3$ ب. $3x$ ج. $3x(1-x)$ د. $3x(1-x^2)$

۶-۱۰. احتمال اینکه دقیقاً یکی از دو پیشامد A و B رخ دهد با کدام گزینه برابر

است؟

الف. $P(A \Delta B)$ ب. $P(A \cup B)$ ج. $P(A \cap B)$ د. $P(A) + P(B)$

۶-۱۱. کدام گزینه نادرست است؟

الف. $P(A \cap B) \leq P(A) + P(B)$ ج. $P(A \cup B) \geq P(A) + P(B)$

ب. $P(A) - P(A \cap B) \leq P(B)$ د. $P(A \cup B) = P(A) + P(B)$

۶-۱۲. اگر $P(A) = ۰٫۶$ و $P(B) = ۰٫۷$ و $P(A|B) = ۰٫۶$ ، آنگاه دو پیشامد هستند.

الف. مستقل ب. وابسته ج. ناسازگار د. متمم
۶-۱۳. اگر $P(A) = ۰٫۶$ و $P(B) = ۰٫۷$ و $P(A|B) = ۰$ ، آنگاه دو پیشامد هستند.

الف. مستقل ب. وابسته ج. ناسازگار د. متمم
۶-۱۴. اگر $P(D|A) = P(D|B) = ۰٫۱$ ، $P(A) = ۰٫۳$ و $P(B) = ۰٫۷$ آنگاه مقدار $P(D)$ برابر کدام گزینه است؟

الف. ۰٫۲ ب. ۰٫۱ ج. ۰٫۳ د. ۰٫۴

۶-۱۵. دستگاهی دارای دو قسمت است که احتمال کارکردن هریک برابر ۰٫۹ است. دستگاه با حداقل یک جزء کار می‌کند. احتمال آنکه دستگاه از کار بیفتد چقدر است؟

الف. ۰٫۸۱ ب. ۰٫۰۱ ج. ۰٫۰۹ د. ۰٫۹۹

فصل هفتم

محاسبات با نرم افزار

هدف کلی

آشنایی با یک نرم افزار آماری که دانشجو را قادر سازد محاسبات پروژه های آماری را در آن محیط انجام دهد.

هدف های یادگیری

خواننده پس از یادگیری این فصل باید بتواند:

- انواع داده ها را وارد نرم افزار نماید.
- دستورات لازم برای ساخت یک جدول آماری اجرا کند.
- انواع نمودارهای آماری را به کمک نرم افزار، ترسیم نماید.
- انواع معیارهای مرکزی (میانگین، میانه، نما، ...)، پراکندگی (دامنه، انحراف معیار، واریانس، ...) و معیارهای ارزیابی فراوانی را که در فصول دوم الی چهارم آمده است به کمک نرم افزار محاسبه نماید.
- دستورات لازم برای محاسبه ی چند سری داده را انجام دهد.
- خروجی های نرم افزار را به محیط های دیگر انتقال دهد.

مقدمه

انتخاب نرم افزار «آر» به عنوان یک ابزار محاسباتی، بیشتر به دلیل نصب رایگان^۱ و آسان و بازبودن منابع آن برای اضافه کردن بسته های نرم افزاری توسط مولفان^۲ است. همچنین قابلیت های به روزرسانی آن که موجب می شود با آخرین یافته های علم آمار هم گام باشد از جمله برتری های این نرم افزار محسوب می شود. این فصل آموزش کامل نرم افزار را در بر ندارد و آنچه آورده شده است صرفاً در سطحی است که بتواند محاسبات آمار توصیفی را انجام دهد. ابتدا چند دستور ساده ولی پرکاربرد برای ورود اطلاعات را شرح می دهیم و سعی می کنیم در هر بخشی، محاسبات یک فصل را با نرم افزار انجام دهیم.

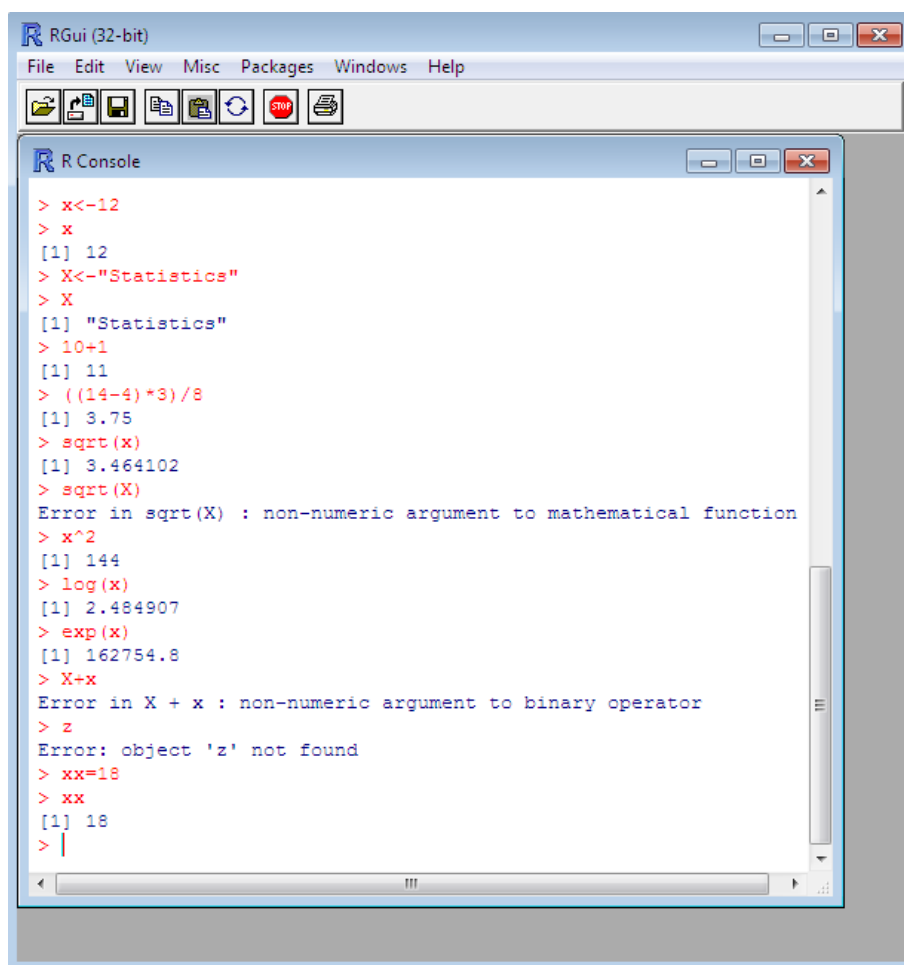
۷-۱ نحوه ی ورود اطلاعات

در این بخش، از دستوراتی استفاده می کنیم که در محیط نرم افزار داده ها وارد می شوند، هرچند قابلیت نرم افزار به گونه ای است که می تواند از سایر محیط های اطلاعاتی داده را بپذیرد و یا اطلاعات را به زبان آنها انتقال دهد.

قاعده ی نام گذاری مقادیر ثابت، متغیرها، بردارها، ماتریس ها و ... این است که اولین کاراکتر، از حروف A-Z یا a-z باشد و مابقی کاراکترها می توانند رقم های صفر الی نه یا «.» و یا زیرخط «_» و یا ترکیبی از همه ی آنها باشند. در محیط «آر»، بین حروف بزرگ و کوچک تمایز قائل می شوند. اگر به جای حرف a در فراخوان، حرف A تایپ شد، نرم افزار پیام خطایی می دهد که این شیء را نمی شناسد. شکل ۷-۱ ضمن آشنایی نمودن خواننده با محیط نرم افزار، چند مثال از تخصیص عدد و کارکتر و همچنین محاسبات جبری را نشان می دهد. قسمت خاکستری که در حاشیه سمت راست ملاحظه می فرمایید، برای نمایش همزمان نمودارها و پنجره های دیگر در محیط نرم افزار است.

^۱ آدرس <http://CRAN.R-project.org>

^۲ open source



```

RGui (32-bit)
File Edit View Misc Packages Windows Help

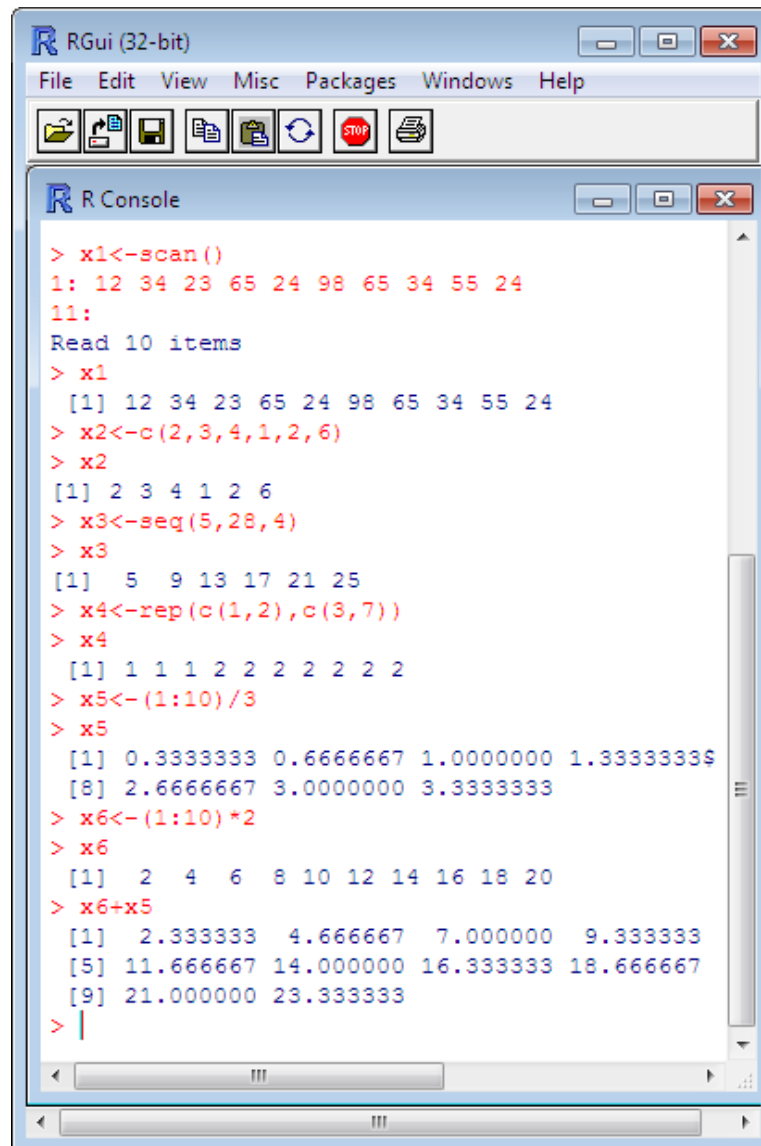
R Console

> x<-12
> x
[1] 12
> X<-"Statistics"
> X
[1] "Statistics"
> 10+1
[1] 11
> ((14-4)*3)/8
[1] 3.75
> sqrt(x)
[1] 3.464102
> sqrt(X)
Error in sqrt(X) : non-numeric argument to mathematical function
> x^2
[1] 144
> log(x)
[1] 2.484907
> exp(x)
[1] 162754.8
> X+x
Error in X + x : non-numeric argument to binary operator
> z
Error: object 'z' not found
> xx=18
> xx
[1] 18
> |

```

شکل ۷-۱. چند مثال از تخصیص یک ثابت، محاسبات جبری در محیط نرم‌افزار

چند تابع مهم را که ورود اطلاعات و ساخت بردار رشته‌ای از اعداد را انجام می‌دهند معرفی می‌کنیم. شکل ۷-۲، همراه با مثال‌هایی توابع `c()`، `scan()`، `rep()`، `seq()` و عملگر «`:`» را نشان می‌دهد. همان‌گونه که در مثال آمده است، عملیات جبری روی این بردارها با دستوراتی که برای حالت ثابت مطرح شد، انجام‌شدنی است.



```

RGui (32-bit)
File Edit View Misc Packages Windows Help

> x1<-scan()
1: 12 34 23 65 24 98 65 34 55 24
11:
Read 10 items
> x1
[1] 12 34 23 65 24 98 65 34 55 24
> x2<-c(2,3,4,1,2,6)
> x2
[1] 2 3 4 1 2 6
> x3<-seq(5,28,4)
> x3
[1] 5 9 13 17 21 25
> x4<-rep(c(1,2),c(3,7))
> x4
[1] 1 1 1 2 2 2 2 2 2 2
> x5<-(1:10)/3
> x5
[1] 0.3333333 0.6666667 1.0000000 1.3333333
[8] 2.6666667 3.0000000 3.3333333
> x6<-(1:10)*2
> x6
[1] 2 4 6 8 10 12 14 16 18 20
> x6+x5
[1] 2.333333 4.666667 7.000000 9.333333
[5] 11.666667 14.000000 16.333333 18.666667
[9] 21.000000 23.333333
>
  
```

شکل ۷-۲. چند مثال مختلف از ورود اطلاعات در محیط نرم‌افزار

۷-۲ ساخت جدول آماری و رسم نمودارها

در این بخش، با دستوراتی آشنا می‌شویم که به‌طور مستقیم اطلاعات وارد محیط نرم‌افزار می‌شود. هر چند که می‌توان به‌کمک دستوراتی اطلاعات از سایر نرم‌افزارها و یا بانک‌های اطلاعاتی به محیط نرم‌افزار آر منتقل کرد.

در شروع فصل دوم در بخش ۲-۱، اولین گام الزامی در سازمان‌دهی داده‌ها و مرتب‌سازی آنها بیان شد. به کمک تابع `sort(.)` داده‌ها را می‌توان از کوچک به بزرگ منظم کرد. (دستور [2]) همچنین برای یافتن بزرگترین و کوچکترین مقدار به ترتیب از توابع `max(.)` و `min(.)` می‌توان استفاده نمود. (دستورات [3] و [4]) محاسبه‌ی فرمول تجربی تعداد طبقات که ترکیبی از عملیات جبری و یک تابع لگاریتمی است در فرمان‌های زیر نوشته شده است. (دستور [5])

```
[1]> x=scan()
44.0 35.4 28.4 37.0 46.0 35.4 19.4 20.4 56.4 43.2
36.2 38.4 49.2 31.8 86.4 12.6 27.4 14.0 39.4 39.4
15.8 28.8 38.0 44.0 38.4 74.0 23.0 11.4 39.8 30.2
29.2 40.6 49.6 30.4 12.2 123.8 42.0 47.0 32.4 39.2
35.2 56.4 31.0 45.0 90.2 100.0 39.0 37.0 49.4 28.2
12.6 27.0 47.8 52.6 41.0 40.0 28.0 23.6 37.6 37.8
30.0 48.5 18.0 41.0 22.6 24.2 89.6 90.4 43.0 29.8
56.2 24.8 12.6 53.6 125.4 16.2 39.0 40.8 33.6 39.4
45.6 37.4 18.0 50.6 103.4 52.4 20.2 64.6 22.2 60.0
42.2 42.0 16.2 108.2 44.0 42.6 39.4 37.6 41.4 40.4
```

```
[2]> y=sort(x)
[3]> xmin=min(x)
[4]> xmax=max(x)
[5]> xk=1+3.3*log(length(x))
```

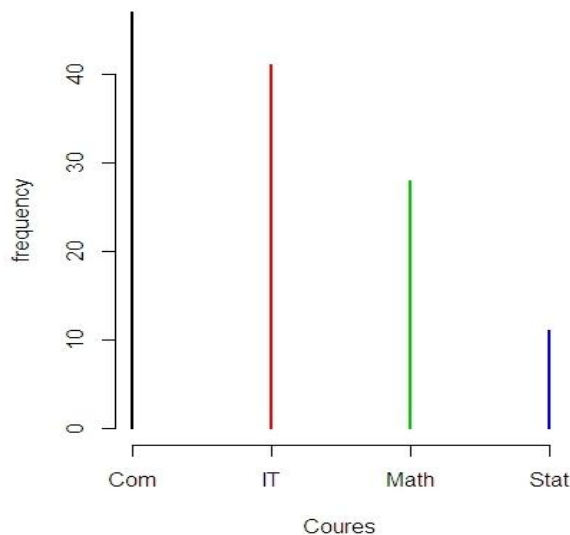
فرض کنید در این مثال می‌خواهید رده‌بندی را در هفت رده انجام دهید. شمارش فراوانی هر رده را می‌توانید از دستوراتی به شرح ذیل انجام دهید.

```
[6]> xl=(xmax-xmin)/7
[7]> xl
[8][1] 16.28571
[9]> f1=sum((((xmin-0.05)<x)&(x>xmin+xl))
[10]> f2=sum((((xmin+xl)<x)&(x>xmin+2*xl))
[11]> f3=sum((((xmin+2*xl)<x)&(x>xmin+3*xl))
[12]> f4=sum((((xmin+3*xl)<x)&(x>xmin+4*xl))
[13]> f5=sum((((xmin+4*xl)<x)&(x>xmin+5*xl))
[14]> f6=sum((((xmin+5*xl)<x)&(x>xmin+6*xl))
[15]> f6=sum((((xmin+6*xl)<x))
[16]> f6=sum((((xmin+5*xl)<x)&(x>xmin+6*xl))
[17]> f7=sum((((xmin+6*xl)<x))
[18]> c(f1,f2,f3,f4,f5,f6,f7)
[19][1] 78 31 11 9 5 2 2
```

اگرچه ممکن است نوشتن این محاسبات از راه‌های دیگری نیز قابل انجام باشد ولیکن برای آموزش جداول آماری به نظر می‌رسد این شیوه موثرتر باشد.

حال می‌خواهیم اطلاعات جدول ۲-۴ را وارد نرم‌افزار کنیم و نمودار مناسبی برای آنها رسم کنیم. طبق دستورات زیر داده‌ها با تابع `rep(.)` وارد می‌شوند و با دستور `table(.)` ساختار یک جدول را می‌یابند، و در فرمان سوم با دستور `plot(.)` نمودار را برای جدول ۲-۴ رسم می‌کنند.

```
> xdata=rep(c("IT","Com","Math","Stat"),c(41,47,28,11))
> table(xdata)
xdata
  Com  IT  Math  Stat
  47  41   28   11
> plot(table(xdata),xlab="Coures",ylab="frequency",col=c(1,2,3,4))
```

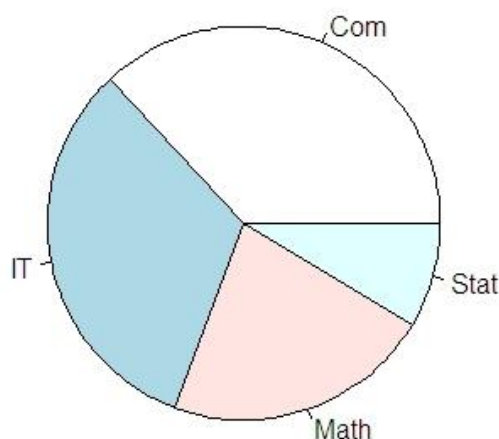


نمودار ۷-۱. نمودار میله‌ای داده‌های جدول ۲-۴

چون داده‌های جدول ۲-۴ از نوع عددی نیستند، نمودار دایره‌ای نیز برای توصیف آنها به کار گرفته می‌شود. با دستور `pie(.)`، به شرح ذیل، می‌توان نمودار دایره‌ای نیز ترسیم کرد. در نمودار ۷-۲ نمودار ساده‌ای از نمودار دایره‌ای رسم شده است که می‌توان با استفاده از سایر شناسه‌های دستور `pie(.)`، تغییر رنگ‌ها، نمایش درصدها، نمایش

فراوانی‌ها، ... را به نمودار افزود.

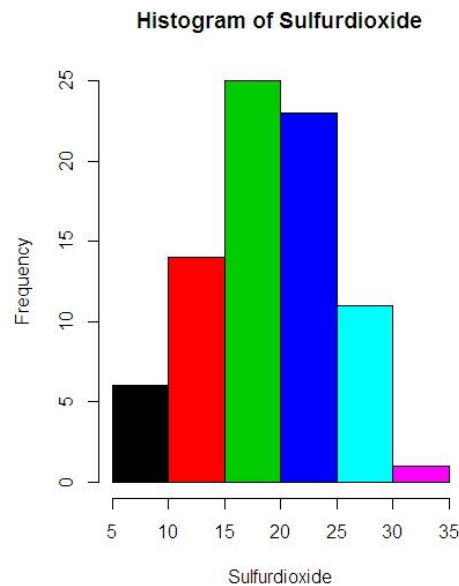
```
>pie(table(xdata))
```



نمودار ۷-۲. نمودار دایره‌ای اطلاعات جدول ۲-۱

در فصل دوم ملاحظه شد که داده‌های مربوط به اکسید سولفور منشره در هوا (جدول ۲-۵) طی مراحل تبدیل به جدول آماری ۲-۷ شد. آنچه در دستورات این بخش آمده است به خوبی می‌تواند کلیه محاسبات دستی را با نرم افزار انجام دهد. حال به رسم نمودار بافت‌نگار به کمک نرم افزار می‌پردازیم. داده‌ها، مطابق شرح در ذیل، با دستور `scan(.)` وارد شده‌اند و دستور `hist(.)` ترسیم نمودار ۷-۳ را انجام داده‌اند. شناسه‌ی `col` در دستور `hist(.)` منجر به نمایش مستطیل‌ها در شش رنگ مختلف شده است.

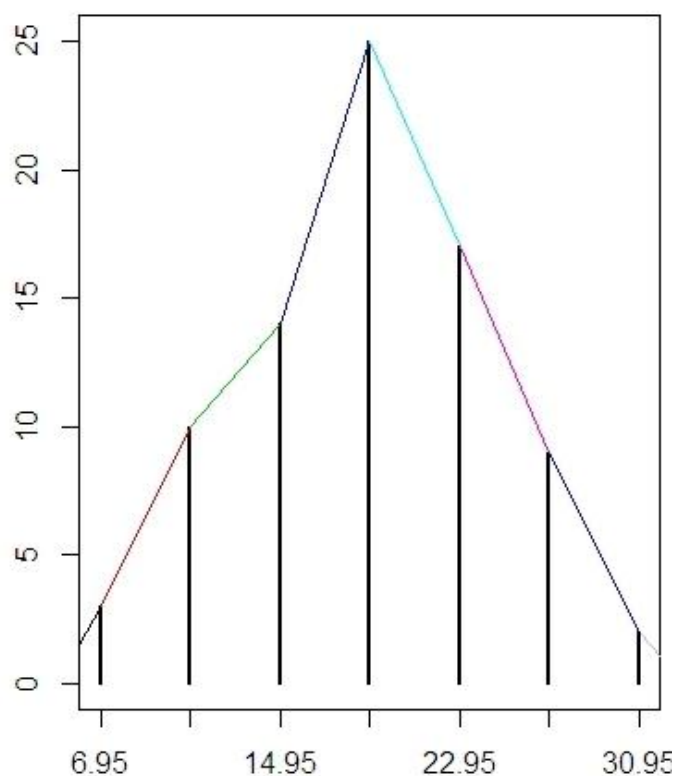
```
> Sulfurdioxide=scan()
1: 15.8 26.4 17.3 11.2 23.9 24.8 18.7 13.9 9.0 13.2
11: 22.7 9.8 6.2 14.7 17.5 26.1 12.8 28.6 28.6 23.7
21: 26.8 22.7 18.0 20.5 11.0 20.9 15.5 19.4 16.7 10.7
31: 19.1 15.2 22.9 26.6 20.4 21.4 19.2 25.6 16.9 19.0
41: 18.5 23.0 24.6 20.1 16.2 18.0 7.7 13.5 23.5 14.5
51: 14.4 29.0 19.4 17.1 20.8 24.3 22.5 24.6 18.4 18.1
61: 8.3 21.9 12.3 22.3 13.3 11.8 19.3 20.0 24.7 31.8
71: 25.9 10.5 15.9 27.5 18.1 17.9 9.4 24.1 20.1 28.5
81:
Read 80 items
> hist(Sulfurdioxide,col=1:6)
```



نمودار ۳-۷. بافت‌نگار اطلاعات جدول ۲-۵ به‌وسیله‌ی نرم‌افزار

برای نمودار چندبهر فراوانی می‌توانیم از دستورات زیر استفاده کنیم:
 دستور `plot(.)` که جدول فراوانی را با کران‌های رده‌های از قبل مشخص شده همراه با فراوانی‌هایشان را در به صورت `table(.)` در شناسه‌ی اولش وارد می‌کنیم. در واقع این قسمت از نمودار، حالت نمودار میله‌ای را ترسیم می‌کند. در ادامه با دستورات `segments(.)` پاره‌خط‌ها روی نمودار بافت‌نگار ترسیم می‌شود. به‌هنگام انجام دستورات در نرم‌افزار، نمودار ۷-۴ را به‌صورت پاره‌خط‌هایی در رنگ‌های مختلف در خروجی نرم‌افزار ملاحظه خواهید کرد.

```
>plot(table(rep(c(6.95,10.95,14.95,18.95,22.95,26.95,30.95),c(3,10,14,25,17,9,2))),ylab=" ")
>segments(4.95,0,6.95,3,col=1)
>segments(6.95,3,10.95,10,col=2)
>segments(10.95,10,14.95,14,col=3)
>segments(14.95,14,18.95,25,col=4)
>segments(18.95,25,22.95,17,col=5)
>segments(22.95,17,26.95,9,col=6)
>segments(26.95,9,30.95,2,col=4)
>segments(30.95,2,32.95,0,col=8)
```

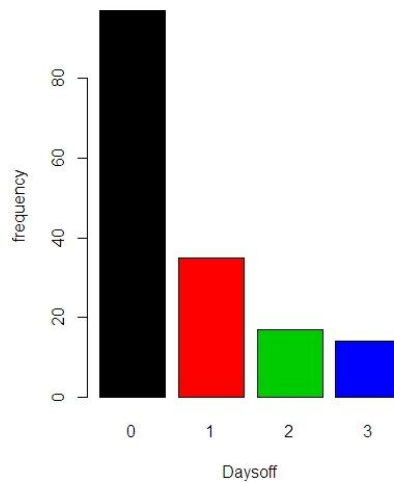


نمودار ۷-۴. نمودار چندبر فراوانی اطلاعات جدول ۲-۵

دستورات مشابه و تکراری را که برای رسم پاره خطها به کار رفته می توان از دستوراتی که برای اجرای حلقه ها مانند for موجود است، نیز انجام داد.

لازم به ذکر است نمودار میله ای که برای اطلاعات جدول ۲-۴، مربوط به مثال ۲-۵ رسم شده است را هم می توانیم مانند عملیاتی که در نمودار ۷-۱ به کار رفته، تکرار کنیم. اگر مایلید نمودار را به صورتی که به ازای هر رده یک مستطیل اختصاص دهد نمایش دهید، آنگاه از دستورات زیر استفاده کنید. جدایی بین مستطیل ها در نمودار ۷-۵ حاکی از گسسته بودن نوع متغیر است.

```
>Daysoff<-rep(0:3,c(97,35,17,14))
>barplot(table(Daysoff),col=1:4)
```



نمودار ۷-۵. نمودار میله‌ای برای اطلاعات مثال ۲-۵

نمودار ساقه‌و برگ را که در فصل دوم بیشتر برای مرتب‌سازی داده‌ها و نمایشی جهت توصیف اطلاعات ارائه شد، می‌توانیم با دستور `stem(.)` انجام دهیم. داده‌های مثال ۲-۲ را در نظر گرفته، و مطابق دستورات در شکل ۷-۳، نمودار ساقه‌و برگ را از نرم‌افزار مشاهده کنید.

```

R Console
> x<-scan()
1: 69 84 52 93 61 74 79 65 88 63 57 64 67 72 74 55 82 61 68 77
21:
Read 20 items
> stem(x)

The decimal point is 1 digit(s) to the right of the |

 5 | 257
 6 | 11345789
 7 | 24479
 8 | 248
 9 | 3
> |

```

شکل ۷-۳. نحوه‌ی انجام یک نمودار ساقه‌و برگ با داده‌های مثال ۲-۲

۷-۳ محاسبه‌ی معیارهای مرکزی و پراکندگی

در نرم افزار «آر» برنامه‌ها و توابع متنوعی وجود دارد که قادر به انجام پروژه‌ها و محاسبات مورد نیاز آماری است. ما در اینجا به توابعی اکتفا می‌کنیم که در محاسبه‌ی معیارهای مرکزی و پراکندگی فصل سوم مطرح شده‌اند. جدول ۷-۹ توابعی را که در محاسبات آمار توصیفی سودمند هستند، نشان می‌دهد.

جدول ۷-۹. برخی توابع پرکاربرد در نرم افزار

تابع	خروجی	مثال
sum(.)	مجموع	> sum(c(1,3,6)) [1] 10
prod(.)	حاصل ضرب	> prod(1:3) [1] 6
mean(.)	میانگین	> mean(c(1,3,6)) [1] 3.333333
median(.)	میانه	> median(c(1,3,6)) [1] 3
quantile(.)	نمایش چند چندک	> quantile(1:99) 0% 25% 50% 75% 100% 1.0 25.5 50.0 74.5 99.0
IQR(.)	انحراف میان چارکی	> IQR(c(2,3,4,2,3,4,6,7)) [1] 1.75
var(.)	واریانس	> var(c(2,3,4,2,3,4,6,7)) [1] 3.267857
sd(.)	انحراف معیار	> sd(c(2,3,4,2,3,4,6,7)) [1] 1.807722

همچنین از تابع **summary(.)** برای نمایش همزمان میانه، چارک اول، چارک سوم، و کوچکترین و بزرگترین مشاهده استفاده می‌کنیم.

```
> summary(c(1,3,6))
Min. 1st Qu. Median Mean 3rd Qu. Max.
1.000 2.000 3.000 3.333 4.500 6.000
```

لازم به ذکر است که برخی محاسبات نسبتاً طولانی‌تر مانند گشتاورهای مرکزی، چولگی و کشیدگی به ترتیب با توابع **moment(.)**، **skewness(.)** و **kurtosis(.)** با

نصب بسته‌ی **e1071** (یا به‌وسیله‌ی فرمان `library(e1071)`) فعال می‌شود. می‌توانیم محاسبه‌ی برخی کمیت‌ها مانند آنچه که در فرمول (۳-۲۱) آمده است، را با یک خط دستور لازم به‌دست آورد. در ادامه یک مثال برای محاسبه‌ی میانگین انحرافات از میانه را آورده‌ایم. تابع `abs(.)` نمایش تابع قدرمطلق است.

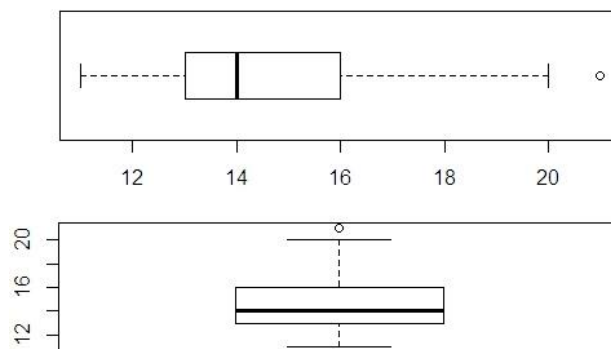
```
> x=c(2,5,3,7,1)
> mean(abs(x-median(x)))
[1] 1.8
```

به‌همین منوال می‌توانیم برای محاسبه‌ی گشتاور مرکزی مرتبه‌ی سوم، و چهارم که در محاسبه‌ی معیارهای چولگی و کشیدگی به‌کار می‌روند، عمل کنیم.

```
> mean((x-mean(x))^3)
[1] 4.032
> mean((x-mean(x))^4)
[1] 37.9712
```

نمودار جعبه-خط که با استفاده از معیارهای مرکزی ترسیم می‌شد، از دستور `boxplot(.)` انجام می‌شود. به مثال زیر که نمودار جعبه-خط در حالت‌های افقی و عمودی، رسم شده‌اند، توجه کنید. (نمودار ۶-۷)

```
> xboxdata=c(11,14,15,16,20,11,14,13,21)
> boxplot(xboxdata,horizontal=TRUE)
> boxplot(xboxdata)
```



نمودار ۶-۷. نمودارهای جعبه-خط در دو حالت افقی و عمودی

خلاصه‌ی فصل هفتم

نرم افزارهای آماری از ابزارهای توانمندی هستند که با آنها محاسبات کمیت‌های عددی و پارامتری را امکان‌پذیر می‌سازد. در این فصل، با دستوراتی برای ورود اطلاعات و ویرایش آنها آشنا شدیم. دستورات لازم برای محاسبات معیارهای مرکزی شامل انواع میانگین‌ها، چندک‌ها، میانه، نما و معیارهای پراکندگی شامل دامنه، انحراف معیار، و واریانس، همراه با مثال‌هایی ارائه شد. دستورات ترسیم نمودارهای آماری نیز مطرح گردید.

تمرین

۷-۱. فرض کنید هدف یافتن پاسخ به این پرسش باشد که «میزان هزینه‌ی استفاده از تلفن همراه در ماه چقدر است؟». بدین منظور جامعه‌ی آماری خودتان را تعریف کنید و سپس اطلاعات یک نمونه ۵۰ تایی را جمع‌آوری نموده و با نرم‌افزار به سوالات زیر پاسخ دهید؟

الف. متوسط سن پاسخ‌دهندگان

ب. وضعیت آخرین مدرک تحصیلی پاسخ‌دهندگان

ج. تعیین کنید چند درصد از پاسخ‌دهندگان زن هستند.

د. متوسط زمان استفاده از تلفن همراه در روز چقدر به دست آورید؟

ه. متوسط هزینه‌ی تلفن همراه در ماه چقدر است؟

و. نمودارهای مناسب برای هریک از یافته‌ها را ترسیم کنید.

۷-۲. مانند تمرین ۷-۱، میزان فاصله‌ی (برحسب کیلومتر) منزل تا دانشگاه را از هریک از هم‌کلاسی‌هایتان جویا شوید و نوع وسیله‌ای که برای رسیدن به دانشگاه از آن استفاده می‌کنند. به کمک معیارهای توصیفی، و نمودارهای آماری، نتایج را به صورت یک گزارش تهیه کنید.

مرجع‌ها

- [۱] مصاحب، غلامحسین (۱۳۵۶). دایرةالمعارف فارسی، دیباچه، جلد اول.
- [۲] بازرگان لاری، عبدالرضا، (۱۳۸۵). مفاهیم پایه‌ای آمار و احتمال، مرکز نشر دانشگاه شیراز.
- [۳] بهبودیان، جواد (۱۳۹۰). آمار و احتمال مقدماتی، آستان قدس رضوی، دانشگاه امام رضا(ع)، ویرایش سوم، چاپ سی و هفتم.
- [۴] ترابی، حمزه (۱۳۸۹). مبانی احتمال، انتشارات تدوین.
- [۵] حسینی پاک، علی اصغر (۱۳۸۶). زمین آمار (ژئواستاتستیک)، انتشارات دانشگاه تهران.
- [۶] عمیدی، علی (۱۳۶۹). احتمال، دانشگاه شهید چمران.
- [۷] گنجعلی، مجتبی، بهرامی سامانی، احسان، (۱۳۸۸). آشنایی با آمار مقدماتی، انتشارات نگارنده دانش.
- [۸] کای‌لای‌چانگ (۱۳۶۴). نظریه‌ی مقدماتی احتمال و فرآیندهای تصادفی، ترجمه‌ی محمد قاسم وحیدی اصل، ابوالقاسم میامی، مرکز نشر دانشگاهی.
- [۹] جان فروند، داندل والپول (۱۳۷۵). آمار ریاضی، ترجمه‌ی دکتر علی عمیدی، دکتر محمد قاسم وحیدی اصل، مرکز نشر دانشگاهی.
- [۱۰] دواس، دی‌ای (۱۳۷۶). پیمایش در تحقیقات اجتماعی، ترجمه‌ی هوشنگ نایی، تهران. نشرنی.
- [۱۱] ساروخانی، باقر (۱۳۷۲). روش‌های تحقیق در علوم اجتماعی، تهران، پژوهشگاه علوم انسانی، جلد اول.

- [12] Ash, R. B. (1972), *Real Analysis and Probability*, New York Academic Press.
- [13] Cleveland, W. S., and McGill, R. (1984), *Graphical Perception: Theory, Experimentation and Application to the Development of Graphical Methods*, Journal of the American Statistical Association, 79, 531–554.
- [14] Kiyosi Ito (1987), *Introduction to Probability Theory*, Cambridge University Press.

- [15] Feller, William (1968), *An Introduction to Probability Theory and Its Applications*, Vols 1 and 2, New York, John Wiley.
- [16] Johnson, N. L. and Kotz, S. (1970), *Continuous Univariate Distributions*, Vols 1 and 2, New York, John Wiley.
- [17] Gordon and Breach (1970); *Elementary Combinatorial Analysis*, Science Publishers Inc, New York.
- [18] Heer, J., and Bostock, M. (2010), *Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design*, in Proceedings of the 28th International Conference on Human Factors in Computing Systems, New York: ACM, CHI 10, pp. 203–212.
- [19] Hoel, P., Port, S. C. and Stone. C. J. (1971), *Introduction to Probability Theory*, Boston, Houghton, Mifflin Co.
- [20] Hogg, R. V. and Craig, A. T., (1973), *Introduction to Mathematical Statistics*, New York, Macmillan Publishing Co.
- [21] Hodges, J. L. and Lehmann, E. L. (1970), *Basic Concepts of Probability and Statistics*, Holden Day San Francisco.
- [22] Majumder, M, Hofmann, H. and Cook, D. (2013), *Validation of Visual Statistical Inference: Applied to Linear Models*, Journal of the American Statistical Association, 108:503, 942-956.
- [23] Ross, S. (1976), *A First Course in Probability*, New York Macmillan Publishing Co.
- [24] Schaeffer, R. L. and Mendenhall (1975), *Introduction to probability Theory and Applications*, North Scituate, Mass Duxbury Press.
- [25] Sevastynov, B. A. (1985), *Problems in the Theory of Probability*, Mir Publishers, Moscow.
- [26] Simkin, D., and Hastie, R. (1987), *An Information Processing Analysis of Graph Perception*, Journal of the American Statistical Association, 82,454–465.
- [27] Tukey, J. W. (1977), *Exploratory Data Analysis*, Boston, MA: Addison-Wesley.

اضافه شود:

پیمایش در تحقیقات اجتماعی / دواس، د.ای/ (۱۳۷۶)، ترجمه هوشنگ نابیی / تهران /
 نشرنی/ چاپ اول
 روش‌های تحقیق در علوم اجتماعی (جلد اول) / دکتر ساروخانی، باقر (۱۳۷۳) تهران /
 پژوهشگاه علوم انسانی/ چاپ اول

پاسخنامه خودآزمایی‌های چهارگزینه‌ای

ج. ۱-۱۶	الف. ۱-۴	الف. ۲-۱۲	ج. ۱-۱
ب. ۱۷-۵	الف. ۲-۴	ج. ۱۳-۲	الف. ۲-۱
ب. ۱۸-۵	د. ۳-۴	ج. ۱۴-۲	ج. ۳-۱
ب. ۱۹-۵	الف. ۴-۴	ب. ۱-۳	د. ۴-۱
ج. ۲۰-۵	الف. ۵-۴	ب. ۲-۳	ج. ۵-۱
الف. ۱-۶	الف. ۶-۴	الف. ۳-۳	د. ۶-۱
الف. ۲-۶	الف. ۱-۵	الف. ۴-۳	الف. ۷-۱
د. ۳-۶	ب. ۲-۵	الف. ۵-۳	ج. ۸-۱
ج. ۴-۶	الف. ۳-۵	ب. ۶-۳	د. ۹-۱
الف. ۵-۶	الف. ۴-۵	الف. ۷-۳	د. ۱۰-۱
ج. ۶-۶	الف. ۵-۵	ج. ۸-۳	الف. ۱-۲
ب. ۷-۶	الف. ۶-۵	ب. ۹-۳	د. ۲-۲
ج. ۸-۶	الف. ۷-۵	د. ۱۰-۳	ب. ۳-۲
الف. ۹-۶	ج. ۸-۵	د. ۱۱-۳	الف. ۴-۲
الف. ۱۰-۶	الف. ۹-۵	ج. ۱۲-۳	ب. ۵-۲
ب. ۱۱-۶	الف. ۱۰-۵	ب. ۱۳-۳	الف. ۶-۲
الف. ۱۲-۶	د. ۱۱-۵	الف. ۱۴-۳	ج. ۷-۲
ج. ۱۳-۶	الف. ۱۲-۵	ج. ۱۵-۳	الف. ۸-۲
ب. ۱۴-۶	الف. ۱۳-۵	ج. ۱۶-۳	الف. ۹-۲
ب. ۱۵-۶	ج. ۱۴-۵	ج. ۱۷-۳	ج. ۱۰-۲
	الف. ۱۵-۵	ج. ۱۸-۳	الف. ۱۱-۲

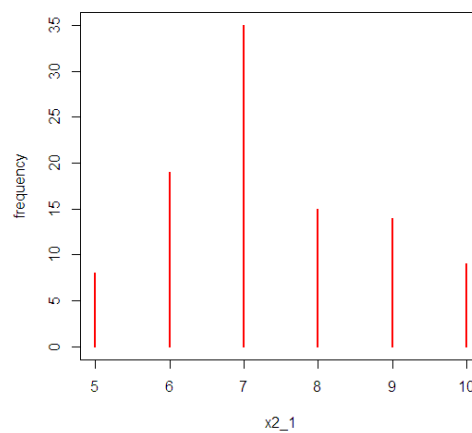
پاسخ تمرین‌ها

- ۱- ۱. جامعه: قطعات خاص تولیدی از ماشین‌سازی تبریز
نمونه: انتخاب تصادفی از قطعات به تعدادی معین
متغیر: طول عمر قطعه؛ متغیر از نوع کمی - پیوسته است.
- ۲- ۱. جامعه: بیماران فشار خون بالا در تهران
متغیر: اندازه‌ی فشار خون
- ۳- ۱. جامعه: دانش‌آموزان ایرانی
نمونه: انتخاب دانش‌آموزان تعلیم‌دیده با روش قدیمی و انتخاب دانش‌آموزان تعلیم‌دیده با روش جدید و مشاهده‌ی وضعیت تحصیلی‌شان
متغیر: میزان یادگیری
- ۴- ۱
- | | |
|------------------------------------|-------------------------|
| عمق: نسبتی - پیوسته | دست‌مزد: نسبتی - پیوسته |
| عرض جغرافیایی: نسبتی - پیوسته | زمان: نسبتی - پیوسته |
| ارتفاع از سطح دریا: نسبتی - پیوسته | جنسیت: اسمی |
| طول جغرافیایی: نسبتی - پیوسته | رضایت از شغل: ترتیبی |
| وزن ماهی: نسبتی - پیوسته | رضایت از درآمد: ترتیبی |
| طول بال پرنده: نسبتی - پیوسته | میزان سواد: ترتیبی |
| | وضعیت تأهل: اسمی |
| | ارتفاع: نسبتی - پیوسته |
- ۲- ۴. الف.

m	f	rf	F
۵	۸	۰٫۰۸	۸
۶	۱۹	۰٫۱۹	۲۷
۷	۳۵	۰٫۳۵	۶۲
۸	۱۵	۰٫۱۵	۷۷
۹	۱۴	۰٫۱۴	۹۱
۱۰	۹	۰٫۰۹	۱۰۰
	۱۰۰	۱٫۰۰	

ب. دستورات برای رسم نمودار به شرح ذیل است.

```
> x2_1<-rep(5:10,c(8,19,35,15,14,9))
> table(x2_1)
x2_1
 5  6  7  8  9 10
 8 19 35 15 14  9
> plot(table(x2_1),ylab="frequency",col=2)
```



۳-۱. ابتدا داده‌ها را مرتب می‌کنیم:

۰ ۰ ۰ ۱ ۱ ۱ ۱ ۲ ۲ ۲ ۳ ۳ ۳ ۳ ۳ ۴ ۴ ۴ ۵ ۶ ۷ ۸

$$Md = \frac{۲+۳}{۲} = ۲٫۵$$

در اینجا دو عدد ۳ و ۱ هر کدام چهار بار تکرار شده‌اند. لذا هر دو مد هستند. برای محاسبه‌ی چارک اول $\frac{n}{4} = \frac{20}{4} = 5$. پس چارک اول پنجمین عدد مرتب‌شده است، یعنی $Q_1 = 1$.

$$\bar{x}_H = \left(\frac{1}{3} \left(\frac{1}{20} + \frac{1}{40} + \frac{1}{80} \right) \right)^{-1} = \left(\frac{1}{3} \times \frac{7}{80} \right)^{-1} = \frac{240}{7} \quad ۲-۳$$

۳-۳. به ازای هر a ای،

$$\begin{aligned} \sum_{i=1}^n (x_i - a)^2 &= \sum_{i=1}^n (x_i - \bar{x} + \bar{x} - a)^2 \\ &= \sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^n (\bar{x} - a)^2 + 2 \sum_{i=1}^n (x_i - \bar{x})(\bar{x} - a). \end{aligned}$$

از طرفی عبارت $(\bar{x} - a)^2$ به اندیس i بستگی ندارد، پس داریم:

$$\sum_{i=1}^n (\bar{x} - a)^2 = n(\bar{x} - a)^2$$

و چون می‌دانیم که $\sum_{i=1}^n (x_i - \bar{x}) = 0$ ، خواهیم داشت:

$$\sum_{i=1}^n (x_i - \bar{x})(\bar{x} - a) = (\bar{x} - a) \sum_{i=1}^n (x_i - \bar{x}) = 0$$

در آخر خواهیم داشت:

$$\sum_{i=1}^n (x_i - a)^2 = \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - a)^2.$$

۳-۴. خیر زیرا

$$2000 < 1854 = 870 + 984 = 5 \times 174 + 8 \times 123 = \text{مجموع وزن افراد}$$

۳-۵.

$$\begin{aligned} \bar{x} &= \frac{575 \times 120000 + 410 \times 117000 + 425 \times 121000}{575 + 410 + 425} \\ &= \frac{168395000}{1410} = 119,429. \end{aligned}$$

۳-۶.

$$\sigma^2 = 2, \quad \bar{x} = 3, \quad \sum_{i=1}^N x_i^2 = 55,$$

$$\sigma^2 = \frac{\sum_{i=1}^n x_i^2 - N \bar{x}^2}{N} = \frac{55 - N \times 3}{N} = \frac{55}{n} - 3 = 2.$$

$$55/n = 5 \rightarrow N = 11.$$

۳-۷. میانگین داده‌های جدید $\bar{x} + 1$ است.

۳-۸. میانگین داده‌های جدید به صورت زیر محاسبه می‌شود:

$$\bar{x}_{new} = \frac{\sum_{i=1}^9 (x_i + i)}{9} = \frac{\sum_{i=1}^9 x_i}{9} + \frac{\sum_{i=1}^9 i}{9} = \bar{x} + \frac{9 \times 10}{9} = 10 + 5 = 15.$$

۳-۹. میانگین داده‌های جدید به صورت زیر محاسبه می‌شود:

$$\bar{x}_{new} = \frac{\sum_{i=1}^{20} (x_i + 2i)}{20} = \frac{\sum_{i=1}^{20} x_i}{20} + \frac{2 \sum_{i=1}^{20} i}{20} = \bar{x} + \frac{2 \times 210}{20} = 15 + 21 = 36.$$

۳-۱۰.

$$\bar{x} = \frac{\sum_{i=1}^{20} x_i}{20} = 15 \rightarrow \sum_{i=1}^{20} x_i = 20 \times 15 = 300,$$

$$\bar{x}_{new} = \frac{\sum_{i=1}^{20} x_i + 30 + 35 + 36}{23} = \frac{300 + 30 + 35 + 36}{23} = \frac{401}{23} = 17.43.$$

۳-۱۱. بنابر صفر بودن انحراف معیار داریم: $x_1 = x_2 = x_3 = x_4 = 25$ ، پس

$$\frac{\sum_{i=1}^4 x_i + 50}{5} = \frac{4 \times 25 + 50}{5} = 30.$$

۳-۱۲. تغییری ایجاد نمی‌شود.

۳-۱۳. $x_1 = x_2 = \dots = x_n = 15$. بنابراین داده‌های مرتب‌شده عبارتند از:

۱۷, ۱۹, ۳۱, ۳۱, ..., ۳۱

در نتیجه $Md = ۳۱$.

۳- ۱۴. اعداد جدید مضرب $\frac{1}{4}$ اعداد قدیمی‌اند. لذا واریانس اعداد جدید $\frac{1}{16}$ اعداد قدیمی و در نتیجه انحراف معیار آنها $\frac{1}{4}$ اعداد قدیمی است.

$$\sigma_{ax+b} = |a| \sigma_x.$$

پس انحراف معیار اعداد جدید برابر است با $۲,۴۵ = ۹,۸ \times \frac{1}{4}$.

۳- ۱۵. اعداد افزوده شده به داده‌های اصلی در دو طرف میانه قرار می‌گیرند، پس میانه‌ی داده‌های جدید برابر با $۵,۶$ است.

۴- ۱. الف.

L_U	L*_U*	f_i	m_i	$f_i m_i$	F	$f_i m_i^2$
۴۴,۵ _ ۵۴,۵	۴۴,۲۵ _ ۵۴,۷۵	۵	۴۹,۵	۲۴۷,۵	۵	۱۲۲۵۱
۵۴,۵ _ ۶۴,۵	۵۴,۷۵ _ ۶۴,۷۵	۴۵	۵۹,۵	۲۶۷۷,۵	۵۰	۱۵۹۳۱۱
۶۴,۵ _ ۷۴,۵	۶۴,۷۵ _ ۷۴,۷۵	۴۳	۶۹,۵	۲۹۸۸,۵	۹۳	۲۰۷۷۰۰
۷۴,۵ _ ۸۴,۵	۷۴,۷۵ _ ۸۴,۷۵	۱۹	۷۹,۵	۱۵۱۰,۵	۱۰۲	۱۲۰۰۸۴
۸۴,۵ _ ۹۴,۵	۸۴,۷۵ _ ۹۴,۷۵	۷	۸۹,۵	۶۲۶,۵	۱۱۹	۵۶۰۷۱
۹۴,۵ _ ۱۰۴,۵	۹۴,۷۵ _ ۱۰۴,۷۵	۱	۹۹,۵	۹۹,۵	۱۲۰	۹۹۰۰
		۱۲۰		۸۱۵۰		۵۶۵۳۱۷

$$\bar{X} = \frac{\sum_{i=1}^n f_i m_i}{120} = \frac{8150}{120} = 67,9,$$

$$\tilde{X} = 64,75 + \frac{70-50}{45} \times 10 = 64,75 + 2,22 = 66,97,$$

$$s^2 = \frac{n(\sum_{i=1}^n m_i^2 f_i) - (\sum_{i=1}^n f_i m_i)^2}{n(n-1)} = \frac{120(565317) - (8150)^2}{120(119)} = 99.$$

ب.

$$\frac{1}{8} \times n = 30 \rightarrow Q_1 = 54,75 + \frac{30-50}{45} \times 10 = 54,75 + 5,55 = 60,3,$$

$$\frac{3}{8} \times n = 90 \rightarrow Q_3 = 64,75 + \frac{90-50}{43} \times 10 = 74,05,$$

$$SK_{p_3} = \frac{3(\bar{X} - Md)}{S} = \frac{3(67,9 - 66,27)}{9,94} = 0,28.$$

۴-۲.

$$SK_{p_3} = \frac{3(\bar{X} - Md)}{S} = \frac{3 \times (3,5 - 3,48)}{1,65} = 0,36.$$

حل تمرین‌ها

۵-۱. $A \cup B$.

۵-۲. اعداد ۲، ۴، ۶ و ۸ مضارب عدد ۲ و عدد ۷ مضرب عدد ۷ هستند. لذا مجموعه‌ی ۲، ۴، ۶، ۷ و ۸ مضارب اعداد ۲ یا ۷ هستند.

۵-۳. حاصل ضرب دکارتی $A \times B$.

۵-۴. به $4 \times 3 = 12$ حالت می‌توان با انتخاب یکی از ۴ جفت جوراب و یکی از ۳ جفت کفش، جوراب و کفش پوشید.

۵-۵. $3 \times 4 \times 5 \times 2 = 120$.۵-۶. $32 \times 32 \times 32 \times 10 \times 10 \times 10 \times 10 = 327680000$.

۵-۷. به‌ازای هر نقطه دو حالت، ۰ و ۱ داریم، پس 2^n تابع می‌توان ساخت.

۵- ۹. الف. $9!$

ب. $6! \times 4!$

۵- ۱۰. الف. $4!$

۵- ۱۱. الف. $\frac{10!}{4! \times 3! \times 2! \times 1!}$

۵- ۱۲. $\frac{9!}{4! \times 3! \times 2!}$

۵- ۱۳. $P_r = \frac{5!}{(5-3)!} = 5 \times 4 \times 3 = 60$

۵- ۱۴. $(4-1)! = 3!$

۵- ۱۵. $\frac{(5-1)!}{2!} = \frac{4!}{2} = 12$

۵- ۱۶. $\binom{5}{2} \binom{7}{3} = \frac{5!}{2!3!} \times \frac{7!}{4!3!} = 10 \times 35 = 350$

در حالت دوم، هر دو مرد را که با هم تفاهم ندارند، در یک گروه در نظر گرفته و از آن گروه فقط یک نفر را انتخاب می‌کنیم و از ۵ مرد باقیمانده دو نفر را انتخاب می‌کنیم:

$$\binom{5}{2} \binom{2}{1} \binom{5}{2} = \frac{5!}{2!3!} \times 2 \times \frac{5!}{2!3!} = 200.$$

۵- ۱۷. $\binom{10}{4} = \frac{10!}{4!6!} = 210$

۵- ۱۸. $\frac{4!}{2! \times 1! \times 1!} = 12$

۵- ۱۹. الف. اگر $x=1$ و $y=1$,

$$\sum_{r=0}^n \binom{n}{r} x^r y^{n-r} = \sum_{r=0}^n \binom{n}{r} = (1+1)^n = 2^n.$$

ب. اگر $x=-1$ و $y=1$,

$$\sum_{r=0}^n \binom{n}{r} x^r y^{n-r} = \sum_{r=0}^n \binom{n}{r} (-1)^r 1^{n-r} = \sum_{r=0}^n \binom{n}{r} (-1)^r = (-1+1)^n = 0.$$

ج. اگر $x=a-1$ و $y=1$,

$$\sum_{r=0}^n \binom{n}{r} x^r y^{n-r} = \sum_{r=0}^n \binom{n}{r} (a-1)^r 1^{n-r} = \sum_{r=0}^n \binom{n}{r} (a-1)^r = (a-1+1)^n = a^n.$$

۵- ۲۰. طبق قضیه‌ی ۵- ۸ داریم:

$$\binom{n}{r} = \binom{n-1}{r} + \binom{n-1}{r-1},$$

از قضیه‌ی ۵-۸ به جای $\binom{n-1}{r-1}$ داریم:

$$\binom{n-1}{r-1} = \binom{n-1-1}{r-1} + \binom{n-1-1}{r-1-1}$$

پس می‌توان نوشت:

$$\begin{aligned} \binom{n}{r} &= \binom{n-1}{r} + \binom{n-1}{r-1} \\ &= \binom{n-1}{r} + \binom{n-1-1}{r-1} + \binom{n-1-1}{r-1-1} \\ &= \binom{n-1}{r-1+1} + \binom{n-1-1}{r-1} + \binom{n-1-1}{r-1-1} + \dots \\ &= \sum_{i=1}^{r+1} \binom{n-i}{r-i+1}. \end{aligned}$$

۵-۲۱. طبق قضیه‌ی ۵-۹ داریم:

$$\sum_{r=0}^k \binom{m}{r} \binom{n}{k-r} = \binom{m+n}{k}.$$

حال اگر $m=n$ و $k=n$ ، خواهیم داشت:

$$\binom{2n}{n} = \sum_{r=0}^n \binom{n}{r} \binom{n}{n-r} = \sum_{r=0}^n \binom{n}{r} \binom{n}{r} = \sum_{r=0}^n \binom{n}{r}^2.$$

۵-۲۲.

$$\sum_{r=0}^n \binom{n}{r} x^{n-1} y^r = (x+y)^n \xrightarrow{x=1} \sum_{r=0}^n \binom{n}{r} y^r = (1+y)^n.$$

از طرفین نسبت به y مشتق می‌گیریم، خواهیم داشت:

$$\sum_{r=0}^n r \binom{n}{r} y^{r-1} = n(1+y)^{n-1},$$

حال اگر قرار دهیم $y=1$ ، خواهیم داشت:

$$\sum_{r=0}^n r \binom{n}{r} = n 2^{n-1}.$$

۵- ۲۳. الف.

$$\binom{\frac{1}{2}}{4} = \frac{\frac{1}{2}(\frac{1}{2}-1)(\frac{1}{2}-2)(\frac{1}{2}-3)}{4!} = \frac{-15}{384}.$$

ب.

$$\binom{-3}{3} = \frac{-3(-3-1)(-3-2)}{3!} = \frac{-60}{6} = -10.$$

۵- ۲۴. الف.

$$\binom{-1}{r} = \frac{-1(-2)(-3)\cdots(-1-(r-1))}{r!} = \frac{(-1)^r (1 \times 2 \times 3 \times \cdots \times r)}{r!} = (-1)^r.$$

ب.

$$\begin{aligned} \binom{-n}{r} &= \frac{-n(-n-1)(-n-2)\cdots(-n-r+1)}{r!} \\ &= \frac{(-1)^r (n(n+1)\cdots(n+r-1))}{r!} \\ &= \frac{(-1)^r}{r!} \frac{(n+r-1)!}{(n-1)!} \\ &= (-1)^r \binom{n+r-1}{r}. \end{aligned}$$

$$\frac{8!}{2!3!3!} = 280. \quad ۵- ۲۵.$$

$$2^3 \times 3^2 \times 4^2 \times \frac{9!}{3! \times 2! \times 3! \times 1!} = 4^3 \times 9!. \quad ۵- ۲۶.$$

$$۶- ۱. S = \{HH, HT, TH, TT\}$$

$$۶- ۲. S = \{245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255\}$$

اندازه‌گیری می‌توان فضای نمونه‌ای آن را به صورت $S = [245, 255]$ اعلام کرد.

$$۶- ۳. S = [0, U]$$

۶-۴. $S = \{0, 1, 2, 3, \dots\}$

۶-۵. $S = \{0, 1, 2, 3, \dots\}$

۶-۶.

۶-۷. الف. $A \cup B = \{x \mid 2 < x < 10\}$ ، ب. $A \cap B = \{x \mid 5 < x < 8\}$ ،
 ج. $A \cap B' = \{x \mid 3 < x < 5\}$ ، د.
 $A' \cup B = \{x \mid 0 < x < 3, 5 < x < 10\}$
 ۶-۸.

$\#(B) = 115$ = تعداد ثبت‌نامی‌های درس ریاضی عمومی

$\#(A) = 138$ = تعداد ثبت‌نامی‌های درس آمار

$\#(A \cap B) = 91$ = تعداد ثبت‌نامی‌های هر دو درس

$\#(A \cup B) = \#(A) + \#(B) - \#(AB) = 138 + 115 - 91 = 162$.

تعدادی که در هیچ‌یک از دروس ثبت‌نام نکرده‌اند

$\#(A \cup B)' = 200 - 162 = 38$.

۶-۹. الف. خودروهایی که به تعمیر موتور، جعبه‌دنده و تعویض تایر نیاز دارند.

ب. خودروهایی که به تایرهای نو موتور و جعبه‌دنده نیاز دارند ولی نیازی به تعمیر کامل موتور ندارند.

ج. خودروهایی که فقط به تعمیر کامل موتور نیاز دارند.

د. خودروهایی که به تعمیر کامل موتور و تایرهای نو نیاز دارند.

ت. خودروهایی که فقط به تعمیر کامل موتور و جعبه‌دنده نیاز دارند یا خودروهایی که فقط به تعمیر جعبه‌دنده نیاز دارند.

و. خودروهایی که فقط به تعمیر جعبه‌دنده یا فقط به تعویض تایر و یا فقط به تعمیر جعبه‌دنده و تعویض تایر نیاز دارند.

۶-۱۰. الف. ناحیه‌ی ۵ ب. ناحیه‌ی ۲

ج. ناحیه‌های ۳، ۵، ۶ د. ناحیه‌ی ۶

۶-۱۱.

$$\begin{aligned}
 P(A \cup B) &= P(A) + P(B) - P(A \cap B) = 0.7 + 0.6 - P(A \cap B) = 0.7, \\
 P(A \cap B) &= 0.1, \\
 P(A' \cap B) &= P(B) - P(A \cap B) = 0.6 - 0.1 = 0.5, \\
 P(A' \cap B') &= 1 - P(A \cup B) = 1 - 0.7 = 0.3, \\
 P(A' \cup B') &= 1 - P(A \cap B) = 1 - 0.1 = 0.9.
 \end{aligned}$$

۶-۱۲.

$$P(A \Delta B) = (A \cap B') \cup (B \cap A') = P(A - B) + P(B - A).$$

۶-۱۳. الف.

$$\begin{aligned}
 P(A \cap B), \quad i &= 2 \\
 P(A \Delta B), \quad i &= 1 \\
 1 - P(A \cup B), \quad i &= 0
 \end{aligned}$$

ب.

$$\begin{aligned}
 1 - P(A \Delta B), \quad i &= 2 \\
 P(A \cup B), \quad i &= 1 \\
 1 - P(A \cup B), \quad i &= 0
 \end{aligned}$$

ج.

$$\begin{aligned}
 1 - P(A \cup B) + P(A \Delta B), \quad i &= 1 \\
 1 - P(A \cup B), \quad i &= 0 \\
 P(A \cap B) + P(A \Delta B) + 1 - P(A \cup B), \quad i &= 2
 \end{aligned}$$

۶-۱۴.

$$\begin{aligned}
 A &= \{2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24\}, \quad B = \{3, 6, 9, 12, 15, 18, 21, 24\} \\
 \#(A \cap B) &= 4
 \end{aligned}$$

$$P(A \cap B) = \frac{4}{24} = \frac{1}{6}.$$

۶-۱۵. الف.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

پس

$$P(A \cap B) = P(A) + P(B) - P(A \cup B) \leq P(A) + P(B),$$

یا

$$P(A \cap B) \leq P(A) + P(B).$$

ب.

$$\begin{aligned} A \cup B &\subseteq S \\ P(S) &\geq P(A \cup B) = P(A) + P(B) - P(A \cap B) \\ 1 &\geq P(A) + P(B) - P(A \cap B) \\ P(A \cap B) &\geq P(A) + P(B) - 1. \end{aligned}$$

۱۶-۶. برای $n=1$

$$P(E_1) \leq P(E_1).$$

فرض می‌کنیم برای $n=k$ برقرار است پس برای $n=k+1$ ثابت می‌کنیم:

$$\begin{aligned} P\left(\bigcup_{i=1}^{k+1} E_i\right) &= P\left(\bigcup_{i=1}^k E_i \cup E_{k+1}\right) \\ &= P\left(\bigcup_{i=1}^k E_i\right) + P(E_{k+1}) - P\left(E_{k+1} \cap \bigcup_{i=1}^k E_i\right) \\ &\leq \sum_{i=1}^k P(E_i) + P(E_{k+1}) = \sum_{i=1}^{k+1} P(E_i). \end{aligned}$$

۱۷-۶. $\frac{P(A)}{1-P(A)} = \frac{a}{b}$ پس

$$bP(A) = a - P(A)$$

$$(a+b)P(A) = a$$

$$P = P(A) = \frac{a}{a+b}$$

۱۸-۶. الف. چون هیچ‌گاه مقدار احتمال منفی نیست.

ب. چون مجموع احتمال‌ها روی کلیه‌ی موارد از یک بیشتر است.

$$0.12 + 0.25 + 0.36 + 0.14 + 0.09 + 0.07 = 1.03 > 1$$

$$0.08 + 0.21 + 0.29 + 0.40 = 0.98 \neq 1$$

ج.

۱۹-۶.

ب. $0.29 + 0.34 + 0.17 = 0.8$

الف. $0.17 + 0.12 = 0.29$

د. $0.08 + 0.29 + 0.34 = 0.71$

ج. $0.34 + 0.17 + 0.12 = 0.63$

۶-۲۰.

$$\begin{array}{ll} \text{الف.} & 0.24 + 0.22 = 0.46 \\ \text{ب.} & 0.15 + 0.03 + 0.22 = 0.4 \\ \text{ج.} & 0.03 + 0.08 = 0.11 \\ \text{د.} & 0.15 + 0.03 + 0.28 + 0.22 = 0.68 \end{array}$$

۶-۲۱.

$$\begin{array}{ll} \text{الف.} & \frac{10+20}{80} = \frac{30}{80} \\ \text{ب.} & \frac{4 \times 5}{80} = \frac{1}{4} \\ \text{ج.} & \frac{2 \times 4}{80} = \frac{1}{10} \\ \text{د.} & \frac{4+2+1+1}{80} = \frac{1}{10} \\ \text{ه.} & \frac{8+14}{40} = \frac{22}{40} = \frac{11}{20} \end{array}$$

۶-۲۲. اگر غیبت دستیار مسن را با A و جوان را با B نشان دهیم، داریم:

$$\begin{array}{ll} \text{الف.} & P(A' \cup B') = P(A') + P(B') - P(A' \cap B') = 0.92 + 0.95 - 0.98 = 0.95 \\ \text{ب.} & P(A \cup B) = P(A) + P(B) - P(AB) = 0.08 + 0.05 - 0.02 = 0.11 \\ \text{ج.} & P(A \Delta B) = P(A) + P(B) - 2P(A \cap B) = 0.08 + 0.05 - 2 \times 0.02 = 0.09 \end{array}$$

$$\text{۶-۲۳. الف. برای هر } A, P(A|B) = \frac{P(A \cap B)}{P(B)} \geq 0.$$

$$\text{ب. } P(B|B) = \frac{P(B \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

ج.

$$\begin{aligned} \frac{P((A_1 \cup A_2 \cup \dots) \cap B)}{P(B)} &= \frac{P((A_1 \cap B) \cup (A_2 \cap B) \cup \dots)}{P(B)} \\ &= \frac{P(A_1 \cap B)}{P(B)} + \frac{P(A_2 \cap B)}{P(B)} + \dots \\ &= P(A_1|B) + P(A_2|B) + \dots \end{aligned}$$

۶-۲۴. الف. $B = \{3, 4, 5\}$, $A' = \{4, 5, 6\}$, $A = \{1, 2, 3\}$, $S = \{1, 2, 3, 4, 5, 6\}$

$$P(B|A) + P(B|A') = \frac{\frac{1}{6}}{\frac{3}{6}} + \frac{\frac{2}{6}}{\frac{3}{6}} = \frac{1}{3} + \frac{2}{3} = 1.$$

ب. بدون تغییر دادن A و S ، قرار دهید $B = \{3\}$ در این صورت:

$$P(B|A) + P(B|A') = \frac{1}{3} + 0 = \frac{1}{3}.$$

۶- ۲۵.

$$\begin{aligned} P(A)P(A|B)P(C|A \cap B) &= P(A) \frac{P(A \cap B)}{P(A)} \frac{P(A \cap B \cap C)}{P(A \cap B)} \\ &= P(A \cap B \cap C). \end{aligned}$$

۶- ۲۶.

$$\begin{aligned} \frac{P(B \cap C)}{P(B)} &= P(C|B) = P(C|A \cap B) = \frac{P(A \cap B \cap C)}{P(A \cap B)} \\ \frac{P(B \cap C)}{P(B)} &= \frac{P(A \cap B \cap C)}{P(A \cap B)} \\ P(A|B \cap C) &= \frac{P(A \cap B \cap C)}{P(B \cap C)} = \frac{P(A \cap B)}{P(B)} = P(A|B). \end{aligned}$$

۶- ۲۷.

$$\begin{aligned} P(T') &= \frac{36 + 27}{90} = \frac{63}{90} \quad \text{ب.} & P(G) &= \frac{18 + 36}{90} = \frac{54}{90} \quad \text{الف.} \\ P(G' \cap T') &= \frac{27}{90} \quad \text{د.} & P(G \cap T) &= \frac{18}{90} \quad \text{ج.} \\ P(G'|T') &= \frac{27}{63} \quad \text{و.} & P(T|G) &= \frac{18}{54} \quad \text{ه.} \end{aligned}$$

۶- ۲۸.

$$P(T|G) \frac{P(T \cap G)}{P(G)} = \frac{\frac{18}{90}}{\frac{54}{90}} = \frac{18}{54}, \quad P(G'|T') = \frac{P(T' \cap G')}{P(T')} = \frac{\frac{27}{90}}{\frac{63}{90}} = \frac{27}{63}$$

۶- ۲۹. اگر $P(B|A) = P(B)$ ، آنگاه داریم:

$$\frac{P(A \cap B)}{P(A)} = P(B) \Rightarrow P(A \cap B) = P(A)P(B) \quad (1)$$

حال

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \xrightarrow{(1)} \frac{P(A)P(B)}{P(B)} = P(A).$$

۶-۳۱. الف.

$$\begin{aligned} & P(\text{از چهارتا انتخاب سه تا سفید باشد} \mid \text{مهره‌ی دوم و سوم سفید}) \\ &= P(\text{از دو مهره‌ی اول یکی سفید و یکی سیاه باشد} \mid \text{مهره‌ی دوم و سوم سفید باشد}) \\ &= \frac{7}{10} \times \frac{6}{9}. \end{aligned}$$

ب.

$$\begin{aligned} & P(\text{اولی سفید باشد}) = \frac{8}{12}. \\ & \text{۶-۳۲. } A: \text{انتخاب از جعبه‌ی اول سفید} \end{aligned}$$

$$\begin{aligned} & P(A \mid \text{از سه جعبه‌ی دو تا سفید انتخاب شده باشد}) \\ &= P(A \mid \text{انتخاب از جعبه‌ی دوم سفید و انتخاب از جعبه‌ی سوم سیاه} \mid A) \\ &+ P(A \mid \text{انتخاب از جعبه‌ی دوم سیاه و انتخاب از جعبه‌ی سوم سفید} \mid A) \\ &= \frac{3}{8} \times \frac{6}{10} \times \frac{4}{6} + \frac{3}{8} \times \frac{4}{10} \times \frac{2}{6}. \end{aligned}$$

توجه داشته باشید که انتخاب جعبه‌ها از هم مستقل هستند.

$$\begin{aligned} & \text{۶-۳۳. از استقلال } A \text{ و } B \text{ داریم } P(A \cap B) = P(A)P(B) \text{ یا} \\ & P(A \mid B) = P(A). \end{aligned}$$

الف.

$$\begin{aligned} P(A' \mid B) &= \frac{P(A' \cap B)}{P(B)} = \frac{P(B - A)}{P(B)} \\ &= \frac{P(B) - P(A \cap B)}{P(B)} = \frac{P(B)}{P(B)} - \frac{P(A \cap B)}{P(B)} \\ &= 1 - \frac{P(A)P(B)}{P(B)} = 1 - P(A) = P(A'). \end{aligned}$$

ب.

$$\begin{aligned}
 P(A'|B') &= \frac{P(A' \cap B')}{P(B')} = \frac{1 - P(A \cup B)}{P(B')} = \frac{1 - P(A) - P(B) - P(A \cap B)}{P(B')} \\
 &= \frac{1 - P(A) - P(B) - P(A)P(B)}{P(B')} = \frac{P(B') - P(A)(1 - P(B))}{P(B')} \\
 &= \frac{P(B')(1 - P(A))}{P(B')} = 1 - P(A) = P(A').
 \end{aligned}$$

۶-۳۵. مطابق شکل ۶-۶ داریم: $P(A) = 0.6$, $P(B) = 0.8$ و $P(C) = 0.5$. پس

$$P(A \cap B \cap C) = 0.24 = 0.6 \times 0.8 \times 0.5 = P(A)P(B)P(C)$$

$$P(A \cap B) = 0.48 = 0.6 \times 0.8 = P(A)P(B)$$

A از B مستقل است،

$$P(A \cap C) = 0.3 = 0.6 \times 0.5 = P(A)P(C)$$

A از C مستقل است، و

$$P(B \cap C) = 0.38 \neq 0.8 \times 0.6 = P(B)P(C) = 0.48$$

B از C مستقل نیست.

ولی در کل هر سه دویه دو مستقل نیستند.

۶-۳۶. در حل تمرین ۶-۳۵ مشاهده می‌کنیم که A از B مستقل است ولی B از C مستقل نیست.

۶-۳۷. در تمرین ۶-۳۵ نشان داده شد که A از B و A از C مستقل است اما

$$P(A \cap (B \cup C)) = 0.552 \neq 0.6 \times 0.92 = P(A)P(B \cup C).$$

پس A از $B \cup C$ مستقل نیستند.

۶-۳۸. وقتی که سه پیشامد A ، B و C مستقل هستند پس:

$$P(A \cap B) = P(A)P(B) \quad (۱)$$

$$P(A \cap C) = P(A)P(C) \quad (۲)$$

$$P(B \cap C) = P(B)P(C) \quad (۳)$$

$$P(A \cap B \cap C) = P(A)P(B)P(C) \quad (۴)$$

الف.

$$\begin{aligned} P(A \cap (B \cap C)) &= P(A \cap B \cap C) \\ &\xrightarrow{(۱)} = P(A)P(B)P(C) \\ &\xrightarrow{(۲)} = P(B)P(B \cap C). \end{aligned}$$

ب.

$$\begin{aligned} P(A \cap (B \cup C)) &= P((A \cap B) \cup (A \cap C)) \\ &= P(A \cap B) + P(A \cap C) - P(A \cap B \cap C) \\ &= P(A)P(B) + P(A)P(C) - P(A)P(B)P(C) \\ &= P(A)(P(B) + P(C) - P(B \cap C)) \\ &= P(A)P(B \cup C). \end{aligned}$$

۶-۳۹. تعداد شرط‌هایی که دو پیشامد در آن‌ها هستند $\binom{k}{۲}$

تعداد شرط‌هایی که سه پیشامد در آن‌ها هستند $\binom{k}{۳}$

$\vdots$

تعداد شرط‌هایی که k پیشامد در آن‌ها هستند $\binom{k}{k}$

پس جواب $\sum_{i=۲}^k \binom{k}{i} = M$ است که حاصل آن برابر با

$$\sum_{i=۰}^k \binom{k}{i} = M + \binom{k}{۰} + \binom{k}{۱} \Rightarrow ۲^k - ۱ - k = M.$$

۶-۴۰. $P(\{T\}) = ۰٫۴۸$ ، $P(\{H\}) = ۰٫۵۲$

الف. $P(\{HHH\}) = ۰٫۵۲ \times ۰٫۵۲ \times ۰٫۵۲ = ۰٫۱۴$

ب. $P(\{TTH\}) = ۰٫۴۸ \times ۰٫۴۸ \times ۰٫۵۲ = ۰٫۱۲۰$

۶-۴۱. $\frac{\binom{۵}{۳}}{\binom{۱۰}{۳}} = \frac{۵ \times ۴ \times ۳}{۱۰ \times ۹ \times ۸} = \frac{۱}{۱۲}$

۶-۴۲. $\frac{۶}{۱۵} \times \frac{۵}{۱۴} \times \frac{۴}{۱۳} \times \frac{۳}{۱۲} = \frac{۳۶۰}{۳۲۷۶۰} = \frac{۱}{۹۱}$

۶-۴۳. الف. $0.9 \times 0.9 \times 0.9 = 0.729$

ب. $0.4 \times 0.1 \times 0.4 + 0.4 \times 0.6 \times 0.4$

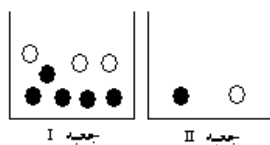
۶-۴۴. پیشامد معیوب بودن E و A_1, A_2, A_3 به ترتیب خطوط تولیدی ۱ و ۲ و ۳ است.

$$\begin{aligned} P(A_1) &= 0.2, & P(A_2) &= 0.3, & P(A_3) &= 0.5, \\ P(E | A_1) &= 0.2, & P(E | A_2) &= 0.3, & P(E | A_3) &= 0.4, \\ P(E) &= \sum_{i=1}^3 P(A_i)P(E | A_i) = 0.5 \times 0.2 + 0.3 \times 0.3 + 0.2 \times 0.4 = 0.27. \end{aligned}$$

۶-۴۵.

$$P(E | A_1) = \frac{P(A_1)P(E | A_1)}{P(E)} = \frac{0.2 \times 0.2}{0.27} = \frac{0.04}{0.27} = \frac{4}{27}.$$

۶-۴۶.

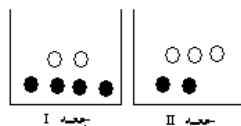


A : انتخاب مهره‌ی سفید از جعبه I $P(A) = \frac{3}{8}$
 A' : انتخاب مهره‌ی سفید از جعبه I $P(A') = \frac{5}{8}$
 B : انتخاب مهره‌ی سفید از جعبه I $P(B | A) = \frac{3}{8}$
 B' : انتخاب مهره‌ی سفید از جعبه I $P(B | A') = \frac{3}{8}$

الف. $P(B) = P(A)(B | A) + P(A')P(B | A') = \frac{3}{8} \times \frac{3}{8} + \frac{5}{8} \times \frac{1}{8} = \frac{11}{32}$

ب. $P(A | B) = \frac{P(A).P(B | A)}{P(B)} = \frac{\frac{3}{8} \times \frac{3}{8}}{\frac{11}{32}} = \frac{6}{11}$

۶-۴۷. الف.



A_1 : پیشامد انتخاب جعبه‌ی I $P(A_1) = \frac{1}{3}$
 A_2 : پیشامد انتخاب جعبه‌ی II $P(A_2) = \frac{1}{3}$

B : پیشامد انتخاب دو مهره‌ی سفید

احتمال انتخاب هر جعبه $\frac{1}{4}$ است پس

$$\begin{aligned} P(B) &= P(A_1).P(B|A_1) + P(A_2).P(B|A_2) \\ &= \frac{1}{2} \times \frac{2}{6} \times \frac{1}{5} + \frac{1}{2} \times \frac{3}{5} \times \frac{2}{4} = \frac{2}{60} + \frac{6}{40} = \frac{22}{120}. \end{aligned}$$

ب.

$$P(A_1|B) = \frac{P(A_1).P(B|A_1)}{P(B)} = \frac{\frac{1}{60}}{\frac{22}{120}} = \frac{18}{22}.$$