

نقش آماری در زندگی روزانه

محمدرضا مشکانی
دانشگاه شهید بهشتی

مقدمه

۱۰ است. هم‌چنین، اتفاق می‌افتد که به مهمانی بروید یا با دوستانش گردهم آمده باشند و بیش از هر روز چای بنوشد. چنین پدیده‌ای را از کودکی تجربه کرده است. پذیرش این تغییرات و نگاه به متوسط برای بیان عادت خود، جزو سرشت او شده است. انحراف از متوسط نیز محدوده‌ی تغییرات عمده را نشان می‌دهد.

با قدری تسامح، این نوع تفکر را در هر زمینه‌ای از اشتغالات انسان‌ها می‌توان دید. همان‌طور که دقت در طبیعت و مشاهده‌ی دقیق تغییرات محیط اطراف موجب پیدایش علومی از قبیل گیاه‌شناسی، جانورشناسی، هواشناسی، شیمی و فیزیک شده است، به همان نحو کنجکاوی انسان‌ها در ماهیت تغییرات تصادفی که در امور مختلف مشاهده شده‌اند، علمی را به وجود آورده است که امروزه به نام آمارشناسی معروف است.

آمار چیست؟

اصولاً در علم آمارشناسی، محور اصلی بحث تغییرات تصادفی است. نمونه‌هایی از این تغییرات تصادفی را برمی‌شماریم: میزان بارش در یک منطقه‌ی معین جغرافیایی از سالی به سالی دیگر تغییر می‌کند. وقوع این تغییر حتمی است، اما میزان تغییر یا کم و زیاد شدن آن تصادفی است، بدین معنی که از پیش نمی‌توان گفت نسبت به سال گذشته چقدر کمتر یا زیادتر خواهد بود. تعداد ضربان نبض افراد در روزهای مختلف و در سنین مختلف

اگر به صحبت‌های روزانه‌ی مردم دقت کنیم، می‌بینیم که مردم اکثراً از مفاهیم آماری برای بیان مقصود خود استفاده می‌کنند، بدون آن که ذکری از آمار به میان آورند. مثلاً، بسیار اتفاق می‌افتد که دو نفر که درباره‌ی عادت‌های زندگی خود با هم صحبت می‌کنند، جمله‌هایی به این صورت بر زبان آورند: «من به طور متوسط روزی ۱۰ استکان چای می‌نوشم»، «شما به طور متوسط هر روز چه مدتی در راه هستید تا به محل کارتان برسید؟» «تغییر می‌کند، بعضی روزها که زودتر بیرون می‌آیم، حدود نیم ساعت و بعضی روزها ۴۵ دقیقه طول می‌کشد؛ تقریباً ۹۵ درصد روزها را بین ۲۵ دقیقه تا ۵۰ دقیقه در راه هستم. آن هم بستگی دارد که چه مسیری را انتخاب کنم.»

در همین چند جمله، دو طرف از مفاهیم تغییرات، میانگین، انحراف از میانگین و بازه‌ی اطمینان استفاده کرده‌اند، بدون آن که صریحاً از آن‌ها نام ببرند. می‌توان گفت که تفکر آماری جزو سرشت آدمیان است و آن هم به تجربه برایشان حاصل شده است. انسان‌ها متوجه شده‌اند که در هر پدیده‌ای، قدری تغییرات تصادفی وجود دارد. همین تغییرات که مقدار آن‌ها قابل پیش‌بینی نیست موجب توجه به مفهوم متوسط یا میانگین شده است. گوینده‌ی اول ما که می‌گوید «به‌طور متوسط روزی ۱۰ استکان چای می‌نوشم»، دریافته است که گاهی به واسطه‌ی اشتغال به برخی کارها، دسترسی او به چای کم است و تعداد استکان‌های چایی که می‌تواند بنوشد کمتر از

تغییرات تصادفی بی‌نظم و غیر قابل پیش‌بینی، سرشت ذاتی پدیده‌هایی است که با آن‌ها سرو کار داریم و مستقیماً بر کیفیت زندگی ما اثر می‌گذارند. نقش آمارشناسان، پیدا کردن یک نظم احتمالی در این بی‌نظمی‌هاست که بر پایه‌ی آن بتوان تصمیم‌هایی را گرفت که احتمال مفید بودن آن‌ها بیش از احتمال زیان‌بخش بودن‌شان است

تغییرات میزان بارندگی و احتمال آمدن سیلاب‌های بزرگ، با ارتفاع کم یا زیاد ساخته می‌شوند. یک پرسش اساسی در این کار این است که در حوزه‌ی آبریز معین با توجه به آمار بارندگی سال‌های گذشته، ارتفاع سد را چگونه تعیین می‌کنند. در این کار محاسبه‌ی احتمال سیلاب‌هایی که هر ۵۰ سال یک بار یا صد سال یک بار با شدتی که تا آن زمان سابقه نداشته رخ می‌دهند، عامل اساسی است. این احتمال را آمارشناسان با بررسی سوابق بارندگی‌ها و سیلاب‌ها به دست می‌آورند.

تا این‌جا دریافتیم که تغییرات تصادفی بی‌نظم و غیر قابل پیش‌بینی، سرشت ذاتی پدیده‌هایی است که با آن‌ها سرو کار داریم و مستقیماً بر کیفیت زندگی ما اثر می‌گذارند. نقش آمارشناسان، پیدا کردن یک نظم احتمالی در این بی‌نظمی‌هاست که بر پایه‌ی آن بتوان تصمیم‌هایی را گرفت که احتمال مفید بودن آن‌ها بیش از احتمال زیان‌بخش بودن‌شان است.

در زیر به چند نمونه از کاربردهای آمار اشاره می‌شود. خواننده‌ی فهیم خود درخواهد یافت که دامنه‌ی کاربرد علم آمار از ثری تا به ثریاست. امتحانات تستی چند گزینه‌ای، آشناترین نوع امتحانات برای کسانی است که به مدرسه رفته‌اند و در آزمون‌های مختلف شرکت کرده‌اند. پرسشی که همواره ذهن امتحان‌کنندگان و امتحان‌شوندگان را به خود مشغول می‌کند این است که یک آزمون تستی معین، تا چه حد میزان دانشی را که می‌خواهد اندازه بگیرد واقعاً اندازه می‌گیرد. سؤال دیگر این است که آیا نتیجه‌ی این آزمون برای یک فرد مورد

تغییر می‌کند. برای آدم‌های سالم محدوده‌ای طبیعی وجود دارد. اگر ضربان قلب آن‌ها از این تعداد کمتر یا بیشتر باشد، نشانه‌ی عدم سلامت است. از منظر بحث ما مهم این است که برای یک آدم سالم نیز ضربان تغییر می‌کند.

شرکت‌های هواپیمایی به تجربه دریافته‌اند که همه‌ی کسانی که بلیت مسافرت خریده‌اند، به دلایلی نمی‌توانند به فرودگاه بیایند و سوار هواپیما شوند. در شمار کسانی که حاضر می‌شوند تغییراتی دیده می‌شود، اما نمی‌توان گفت که چند درصد حاضر نخواهند شد. از این رو پس از سال‌ها تجربه بدین نتیجه رسیده‌اند که حدود یکی دو درصد بیشتر بلیت بفروشند تا هواپیما با صندلی خالی پرواز نکند. ممکن است شاهد صحنه‌ای هم بوده باشید که بر خلاف انتظار، همه‌ی مسافران بلیت‌دار حاضر شده‌اند و شرکت هواپیمایی با دادن جایزه‌هایی بعضی مسافران را قانع می‌کند که با پروازهای بعدی سفر کنند. در اینجا نیز نفس تغییرات مهم است. پرسشی مقدر این است که این جایزه‌ها چقدر باید باشند که ضرری به شرکت هواپیمایی وارد نشود. قبل از ورود به بحث اصلی همین‌جا می‌توانیم بگوییم که به کمک محاسبات آماری می‌توان دوگونه عمل کرد: یا درصد فروش اضافی را ثابت گرفت و میزان جایزه‌ها را طوری تعیین کرد که احتمال زیان شرکت خیلی کم باشد، یا برعکس جایزه را ثابت گرفت و متوسط تعداد بلیت‌های اضافی را تعیین کرد.

سدهایی که بر روی رودخانه‌ها بسته می‌شوند نیز به پیروی از

آمارشناسان با بررسی سوابق امر و به کار گرفتن قانون‌های احتمال و روش‌های آماری، احتمال هریک از پیامدهای موردنظر را حساب می‌کنند و در راستای استفاده‌ی صحیح از نتایج، به تصمیم‌گیران کمک می‌کنند

از این خطاها چقدر باشد. پیداست که در این‌جا هم چهار حالت داریم: ۱. بیمار واقعاً به درستی «بیمار» تشخیص داده شده، ۲. بیمار به خطا «سالم» تشخیص داده شده، ۳. سالم به خطا «بیمار» تشخیص داده شده و ۴. سالم به درستی «سالم» تشخیص داده شده است. پیشامد اول و چهارم را «ویژگی» آزمایش و پیشامد دوم و سوم را «حساسیت آزمایش» می‌نامند. روشن است که آزمایش‌هایی که هم ویژگی آن‌ها و هم حساسیت آن‌ها دارای احتمال زیاد باشند، برترند؛ زیرا از خطاهایی که ممکن است زیان‌بخش و غیرقابل جبران باشند، با احتمال زیاد جلوگیری می‌کنند. آمارشناسان به کمک پزشکان با گردآوری و تحلیل داده‌هایی درباره‌ی روش‌هایی که جدیداً معرفی شده‌اند، می‌توانند این احتمال‌ها را حساب کنند.

از قضا این دو مفهوم در امور نظامی نیز کاربردهایی مشابه دارند. مثلاً در حملات هوایی دشمن، دو طرف جنگ با به‌کارگیری رادار، وضعیت حمله‌های هوایی را زیر نظر دارند. در قرائت‌های رادار نیز ممکن است «هشدار» درست یا «هشدار» نادرست داشته باشیم. هشدار درست یعنی رادار نشان می‌دهد که حمله‌ای در شرف وقوع است به شرط آن که چنین حمله‌ای در کار باشد. هشدار نادرست یعنی حمله‌ای در کار نیست و رادار به اشتباه علامت می‌دهد. اما ممکن است حمله‌ای در کار باشد و رادار علامت ندهد. پیداست که در این‌جا احتمال هشدار نادرست، حمله‌ای در شرف وقوع است و رادار علامت نمی‌دهد، مرگ‌آفرین است و رادار باید چنان ساخته شود که احتمال آن هرچه ممکن است کوچک باشد. همین وضعیت در مورد تلفن‌هایی که به ایستگاه‌های آتش‌نشانی می‌رسند دیده می‌شود. در این‌جا نیز اگر آتشی در گرفته باشد و متصدی تلفن فکر کند جزو تلفن‌های مزاحم است، پیامد آن بسیار ناگوار خواهد بود. آمارشناسان با بررسی سوابق امر و به کار گرفتن قانون‌های احتمال و روش‌های آماری، احتمال هریک از پیامدهای موردنظر را حساب می‌کنند و در راستای استفاده‌ی صحیح از نتایج، به تصمیم‌گیران کمک می‌کنند.

در شاخه‌ای از آمار به نام «آمارزیستی»، آمارشناسان بیشتر با داده‌های حاصل از فعالیت‌های پزشکی و زیست‌شناختی سرو کار دارند. در این‌جا یکی دو نمونه از کمک آمارشناسان به پزشکان را ذکر می‌کنیم تا خواننده دریابد که آمار چه کمک‌های ارزشمندی می‌تواند در شاخه‌های مختلف علوم و فنون عرضه کند.

آزمون یا به قول آزمون‌کنندگان برای فردآزمودنی تا چه حد قابل تکرار است. یعنی اگر همین پرسش‌ها مثلاً دو روز دیگر از همین آزمودنی پرسیده شوند، نمونه‌ی امروز او با نمونه‌ی دو روز دیگر او چقدر شباهت خواهد داشت؟

روایی و پایایی

دو مفهومی که در بالا از آن‌ها یاد کردیم، در روان‌سنجی و سنجش و اندازه‌گیری به روایی (اعتبار) و پایایی (قابلیت اعتماد) معروف‌اند. این دو مفهوم در کارهای روزمره بسیار آشنا هستند. مثلاً اگر ترازویی داشته باشیم و اجسام را با آن وزن کنیم آیا برای جسمی معین و وزنی مشخص، ترازو وزنی نزدیک به وزن واقعی به دست خواهد داد. اگر چنین باشد، این ترازو «معتبر» است. آیا اگر توزین را بارها و بارها انجام دهیم، صرف‌نظر از این که متوسط وزن‌های به دست آمده به مقدار واقعی نزدیک است یا نه، این وزن‌ها تغییرات تصادفی اندکی را نشان می‌دهند. اگر چنین باشد این ترازو «قابل اعتماد» است. پس ممکن است ترازویی «معتبر و قابل اعتماد» یا «معتبر و غیر قابل اعتماد»، «نامعتبر و قابل اعتماد»، «نامعتبر و غیر قابل اعتماد» باشد. این حالت‌ها را در آمار با «ادبیی» (تفاوت میانگین وزن‌ها با وزن واقعی جسم) و «واریانس» (متوسط مجذور انحرافات از میانگین) توصیف می‌کنند. در مبحثی از آمار به نام طرح آزمایش‌ها، از این دو ویژگی با صفت‌های «درستی» و «دقت» یاد می‌کنند.

این دو مفهوم شباهت زیادی به ویژگی‌های یک آزمایش پزشکی دارد. در آزمایشگاه‌های پزشکی که برخی ویژگی‌های فیزیولوژیک انسان‌ها را اندازه می‌گیرند؛ مانند اوره‌ی خون، قند خون، کلسترول مضر و مفید، ابتلای به بیماری‌هایی مانند ایدز، سل، هیپاتیت، سرطان پروستات و غیره، پزشکان نگران دو موضوع هستند. اگر آزمایش‌های کلینیکی خاصی برای تشخیص یک بیماری ابداع شده باشد، نخستین پرسش‌هایی که به ذهن پزشکان می‌رسد این است که آیا این آزمایش، فرد واقعاً بیمار را «بیمار» و فرد سالم را «سالم» تشخیص می‌دهد یا امکان دارد، فرد واقعاً سالم را «بیمار» و فرد واقعاً بیمار را «سالم» تشخیص دهد. دو حالت اخیر تشخیص اشتباه هستند و مهم این است که احتمال هریک

می‌توان گفت که تفکر آماری جزو سرشت آدمیان است و آن هم به تجربه برایشان حاصل شده است. انسان‌ها متوجه شده‌اند که در هر پدیده‌ای، قدری تغییرات تصادفی وجود دارد. همین تغییرات که مقدار آن‌ها قابل پیش‌بینی نیست موجب توجه به مفهوم متوسط یا میانگین شده است

مانند دما، بارندگی، شانس ابتلا به بیماری به دلیل استفاده از مواد مضر مانند سیگار و غیره و غیره استفاده کرد. در اقتصاد از این روش استفاده می‌شود و به آن نام اقتصادسنجی داده‌اند. از مثال‌هایی که در بالا ذکر شد می‌توان گفت که آمار نوعی روش حل مسئله در رویارویی با تغییرات تصادفی غیر قابل پیش‌بینی است که از آن به نام **عدم حتمیت** یاد می‌کنند. همه‌گیر بودن تغییرات تصادفی موجب گسترش آمار در رشته‌های بسیار متعدد شده است و در همه‌ی آن‌ها از اصول ثابتی استفاده می‌شود که همان محتوای علم آمار است. مثلاً در حالی که مهندسان به احتمال کارکردن یک دستگاه پس از انجام تعمیر روی آن علاقه‌مندند، پزشکان به احتمال زنده ماندن یک فرد پس از یک جراحی یا معالجه توجه دارند. هر دوی آن‌ها با مفهوم واحدی به نام **تحلیل بقا** سروکار دارند. از این‌گونه تناظرها فراوان وجود دارند که به برخی از آن‌ها در بالا اشاره شد.

مأموریت آمار

آمار مأموریت خود را با طراحی روش گردآوری داده‌ها درباره‌ی مسئله‌ای خاص، خلاصه‌سازی و جدول‌بندی و مرتب کردن داده‌های گردآوری شده، تحلیل داده‌ها برای یافتن قوانین کلی حاکم بر آن‌ها و درک ماهیت تغییرات و صورت‌بندی آن به صورت یک توزیع احتمال که احتمال حالت‌های مختلف را به دست می‌دهد، و نتیجه‌گیری از داده‌های تحلیل‌شده انجام می‌دهد. در این رهگذر، خدمات بسیار برجسته‌ای برای درک و تحلیل پدیده‌هایی که در نگاه اول آشفته و بی‌قانون به نظر می‌رسند انجام داده است. خواننده‌ی علاقه‌مند را برای آگاهی بیشتر از حیطه‌ی عمل آمار و نمونه‌هایی از کاربرد آن در زمینه‌های بسیار بسیار گسترده (یا چنان‌که گفته شد از ثری تا به ثریا) به کتاب «آمار راهی به سوی ناشناخته‌ها» از انتشارات انجمن آمار ایران ارجاع می‌دهیم.

بی‌نوشت

رابطه‌ی بین سرخچه در زنان باردار و نقص‌های مادرزادی کودکان آن‌ها موضوعی نیست که به آسانی بتوان به آن پی برد، زیرا این بیماری می‌تواند نسبتاً خفیف باشد. پس از آن که چشم‌پزشکی به نام نورمان گرگ و دیگران به این پیوند مشکوک شدند، آمارشناسی از اهالی استرالیا به نام اچ. او. لاکاستر توانست نشان دهد که در سرشماری‌های سال‌های ۱۹۱۱، ۱۹۲۱، و ۱۹۳۳ استرالیا در بین کسانی که در همه‌گیری سرخچه در سال ۱۸۹۹ متولد شده بودند، ناشنوایی بیش از متولدین سال‌های دیگر بود. اتفاقاً این شخص کشف دیگری در خصوص رابطه‌ی بین آفتاب و سرطان پوست نیز دارد. وی با بررسی داده‌های دقیق و مقایسه‌ی نرخ ابتلا به سرطان پوست در عرض‌های جغرافیایی مختلف کشور استرالیا نشان داد که نرخ ابتلا به این بیماری با عرض جغرافیایی همبستگی قوی دارد. در ایالت‌های شمالی استرالیا نور آفتاب شدیدتر از ایالت‌های جنوبی است. نتیجه آن که سال‌ها قبل از پیدا شدن حفره‌ی در لایه‌ی اوزون، وی به این رابطه از راه روش‌های آماری پی برد. این بحث را با ذکر مثالی از رشته‌ی زیست‌سنجی که شاخه‌ای از آمار است به پایان خواهیم برد.

والدین بلند قد به طور متوسط، فرزندان کوتاهی از خود دارند. این واقعیت را آمارشناسی به نام فرانسیس گالتون (۱۸۲۲-۱۹۱۱) از اهالی انگلستان به اثبات رساند. او و شاگرد و همکارش کارل پرسون (۱۸۵۷-۱۹۳۶) تعداد ۱۰۷۸ زوج پدر و پسر را مورد بررسی قرار دادند و دریافته‌اند که پدرانی که قدی بلندتر از میانگین دارند، صاحب پسرانی هستند که قدشان نسبت به پدر کوتاه‌تر است. برعکس پدرانی که قدی کوتاه‌تر از میانگین دارند پسرانشان به نسبت بلند قدترند. گالتون این پدیده را «برگشت به میان‌میگی» یا عودت به حالت معمولی نامید. روش آماری این کشف^۱ رفته رفته به رگرسیون تخفیف یافت و سرچشمه‌ی کاربرد آن در زمینه‌های مختلف شد. شاید به جرأت بتوان گفت که امروزه رگرسیون در شکل‌های مختلف آن پرکاربردترین روش آماری است که به صورت تابعی یا مدلی میانگین یک متغیر تصادفی را بر حسب مقدارهای مختلف متغیر (یا متغیرهای) پیشگو بیان می‌کند. از این روش می‌توان برای پیشگویی بهای سهام در بازار بورس و پیش‌بینی پدیده‌هایی که در طی زمان تحول می‌یابند

1. Regression to Mediocrity