

کاربرد آمار در کتابخانه

نوشته استوارت هنبوس^۱
ترجمه فریده عصاره^۲

چکیده: این مقاله تلاش کرده است که سه کاربرد ساده آمار را در کتابخانه‌ها معرفی کند که هر کتابدار بدون داشتن دانش وسیع آماری بتواند آنها را به کار برد. دو روش اول می‌تواند به صورت دستی انجام شود (در مورد همبستگی کار ساده‌تر خواهد بود اگر همبستگی اسپیرمن مورد استفاده واقع شود تا پیرسون که نرم‌افزار مینی‌تب به کار برده است و مربع‌کای با فرمول ویژه‌اش که در خیلی از کتابهای آمار توضیح داده شده است). لزومی ندارد که کتابدار برای انجام این امور خدمات کتابخانه را متوقف کند. خیلی از داده‌ها همواره موجودند یا بدون هیچ‌گونه هزینه‌ای می‌توانند به وجود بیایند. این‌گونه تجزیه و تحلیلها احتمالاً در کتابخانه‌ها انجام می‌شوند. تصمیمهای مطلوب می‌تواند براساس تجربه معقول مبتنی بر داده‌های معتبر صورت گیرد و تحلیلی که از آمار می‌کنیم، در حرکت از این انبوه داده‌های خام که بر حسب وظیفه گردآوری کرده‌ایم، به سوی تصمیم‌گیرهای لازم، نقش مفیدی ایفا می‌کند.

سالهاست که کتابخانه‌ها به جمع‌آوری آمار می‌پردازند. از آن‌جمله آمار مواد به امانت رفته و امانت‌گیرندگان، داده‌هایی درباره سفارشات، دریافتها و بودجه‌های هزینه شده، و نیز اطلاعاتی پیرامون دیرکردها و درخواستهای مراجعان. گردآوری این‌گونه آمارها در کتابخانه‌ها به صورت

1. Stuart Hannabuss. "Statistics at work in the Library" in: *Library Review*, Vol. 41, no. 4, 1992. P 38-48.

۲. عضو هیئت علمی دانشکده علوم تربیتی و روانشناسی دانشگاه شهید چمران (اهواز) و دانشجوی دوره دکتری کتابداری و اطلاع‌رسانی دانشگاه نیوسات ولز - استرالیا.

روزانه امری کاملاً عادی است، زیرا هرگونه تصمیم‌گیری می‌تواند براساس این آمارها صورت گیرد. مثلاً تغییرات منحنی امانات در طول سال می‌تواند نشانگر شکل ویژه میزان کار کارکنان بخش امانت باشد؛ یا تأثیر تورم بر قیمت مجلات نتیجتاً به قطع اشتراک برخی از آنها بیانجامد؛ و همچنین میزان درخواست برای تک‌نگاشتها می‌تواند به خرید بیشتر و امانت گرفتن کمتر کتاب منجر شود. تمام این موارد نمونه‌هایی از کاربرد آمار عملی در کتابخانه‌هاست.

اکثر این آمارها تزایدی است. به این معنی که در جریان کار، گردآوری می‌شوند نظیر: تعداد تک‌نگاشتهای تهیه شده، نوارهای به امانت رفته و کودکانی که از کتابخانه‌ای خاص استفاده می‌کنند، و نیز اغلب به‌صورتی ساده قابل تهیه و درک است. در کتابخانه‌هایی که بخشهای گردش کتاب و سفارش و تهیه مواد به‌صورت ماشینی اداره می‌شود، محاسبه آمارهای فوق به سادگی از طریق برنامه‌ریزی کامپیوتری میسر است. انواع رسم نگارها^۱ نیز برای تهیه این قبیل داده‌ها می‌تواند به کار رود. این‌گونه آمارها می‌تواند برای کتابداران و حسابداران در تمام سطوح اطلاعاتی قابل درک باشد.

به‌هر حال این نوع داده‌ها همواره روندی افزایشی دارند. برای مثال اطلاعات مالی یا ارقام مربوط به امانات، اغلب به‌صورت تجمعی و کلی، در طول سال افزایش می‌یابند، یا به بخشها و قسمتهای کتابخانه تقسیم می‌شوند. آن‌گاه که این آمارها وارد گزارش سالانه شوند ممکن است کاهش یافته باشند، اما اکثراً فقط بیانگر نمایی کلی از وضعیت واقعی کتابخانه هستند.

دانستن آمار کلی مفید است، اما پرداختن زیاد به آمار کلی منجر به غفلت از اهمیت تجزیه و تحلیل دقیقتر ارقام می‌گردد. این آمار، داده‌ها را در مقیاس خرد بررسی می‌کند. در حالی که آمار کلی سطح کلان اقتصاد خدمات کتابخانه را مورد بررسی قرار می‌دهد.

هدف این مقاله آن است که سه مورد از آمارهایی را که در کتابخانه‌ها معمول است بررسی کند، و نشان دهد که چگونه روشهای آماری می‌توانند برای پاسخ‌دهی به سؤالات عملی یا بیان شیوه دستیابی به پاسخ، مورد استفاده قرار گیرند.

نخست، به تحقیق مختصری در زمینه مدت دیرکردها و امانات پرداخته و داده‌ها را با استفاده از همبستگی تجزیه و تحلیل می‌کنیم. دوم، دیرکردهای تعدادی از استفاده‌کنندگان را بررسی می‌کنیم تا مشخص شود آیا یافته‌ها از نظر آماری معنی‌دار هستند یا خیر. سوم، با داشتن داده‌هایی درباره وضعیت جاری امانت، بررسی خواهیم کرد که چگونه می‌توانیم تعیین کنیم که مدارک هر چند وقت یک‌بار مورد استفاده قرار می‌گیرند، و در اینجا از رگرسیون استفاده خواهیم

کرد. تمام این روشها به سادگی قابل درک است. انجام محاسبه آنها به صورت دستی میسر است، اگرچه با استفاده از نرم افزارهای آماری بسیار ساده تر انجام می شود. در این مقاله نرم افزار مینی تب^۱ مورد استفاده واقع شده است.^۲

دیرکردها و مدت امانت

همواره بین دیرکردها و مدت امانت رابطه‌ای وجود دارد. اگر مدت امانت خیلی طولانی باشد خوانندگان وقت کافی برای مطالعه کامل مدرک خواهند داشت، اما ممکن است زمان برگشت کتاب به کتابخانه را فراموش کنند. به عبارت دیگر، مدت طولانی امانت می تواند در برگشت کتاب به کتابخانه تأثیر داشته باشد. اگر مدت زمان امانت کوتاه باشد، خوانندگان تحت فشار قرار می گیرند که سریعتر کتاب را مطالعه کنند، و به کتابخانه برگردانند و این امتیاز را نیز دارد که کتاب میان استفاده کنندگان بیشتری به گردش درآید. از طرف دیگر مدت زمان بسیار کوتاه امانت می تواند موجب تأخیر در برگشت کتاب به کتابخانه شود، زیرا خواننده زمان امانت را برای مطالعه کتاب کافی نمی یابد و در نتیجه کتاب را دیرتر به کتابخانه باز می گرداند. در اینجا ترکیبی از عوامل اداری، روانشناختی، و مالی (اگر جریمه در کار باشد) دخیل هستند.

می توان گفت که میان دیرکردها و مدت زمان امانت نوعی رابطه آماری وجود دارد. برای مثال اگر بتوانیم ثابت کنیم که با کوتاه کردن مدت امانت کتابها از سه هفته به دو هفته میزان دیرکردها ۱۵ درصد کاهش می یابد، در آن صورت می توان گفت که نوعی رابطه آماری میان دیرکردها و مدت زمان امانت وجود دارد، که آماردانان این رابطه را همبستگی^۳ می نامند.

همبستگی می تواند مثبت باشد (وقتی که افزایش یک متغیر به بالا رفتن متغیر دیگر یا کاهش یک متغیر به پایین آمدن متغیر دیگر بیانجامد) یا منفی باشد (وقتی که افزایش یک متغیر به کاهش و کاهش یک متغیر به افزایش متغیر دیگری منجر شود، یعنی متغیرها در جهت مخالف همدیگر حرکت کنند). نمونه روشن همبستگی مثبت در کتابخانه عبارت است از تهیه بیشتر کتاب وقتی که بودجه خرید کتاب افزایش می یابد. نمونه روشن همبستگی منفی در کتابخانه نظیر روش امانت کتابهای رشته کامپیوتر است، که هرچه کهنه تر و قدیمی تر می شوند کمتر به

1. Minitab

۲. توصیه می شود که در صورت عدم دسترسی به نرم افزار فوق الذکر، نرم افزار اکسل (Excel) اعم از ویرایش ۴ یا ۵ آن به کار گرفته شود. این نرم افزار به سهولت کلیه محاسبات این بررسی را انجام و نتایج را ارائه می نماید..

3. Correlation

امانت می‌روند.

رابطه همبستگی میان دو یا چند متغیر می‌تواند قوی یا ضعیف باشد. مانند کهنگی کتاب که می‌تواند بر میزان به امانت رفتن آن تأثیر زیادی نداشته باشد، این مطلب در حوزه علوم انسانی در کتابخانه دانشگاهی مشهود است، در ویرایشهای قدیم و نیز ویرایشهای جدید یک کتاب به میزان وسیعی مورد استفاده دانشجویان قرار می‌گیرد. گاهی ممکن است بین دو متغیر اصلاً همبستگی نباشد، چیزی که همبستگی صفر نامیده می‌شود. نمونه همبستگی صفر ممکن است به این صورت باشد که همزمان با افزایش بودجه کتابخانه، سیستم تهویه کتابخانه چندین مرتبه خراب شود، که در این دو عامل هم از لحاظ آماری و هم در زندگی واقعی هیچ‌گونه رابطه‌ای با هم ندارند.

وجود رابطه میان دو متغیر ضرورتاً دلالت بر رابطه علت و معلولی ندارد؛ مثلاً ممکن است ازدیاد بودجه نشریات ادواری، اشتراک مجلات بیشتر و متنوع‌تر و نیز ازدیاد استفاده کنندگان کتابخانه را دربر داشته باشد. اما ازدیاد استفاده کنندگان کتابخانه در آن زمان خاص می‌تواند متأثر از عوامل دیگری نیز باشد، مثلاً در آن زمان از سال دانشجویان برای تهیه پروژه‌هایشان تعداد زیادی از مجلات را به کار برده باشند یا تغییر جمعیت در آن منطقه از دو سال پیش تا آن زمان باعث افزایش تعداد استفاده کنندگان مجلات شده باشد. برای روشنتر شدن مطلب مثال دیگری می‌آوریم. ممکن است دریابیم که تعداد مالکین سگ در نیویورک و میزان کاشت درخت در کورنوال^۱ هر دو افزایش داشته‌اند، اما جست و جوی رابطه علت و معلولی میان آنها بی‌ربط و نامناسب خواهد بود.

به مثالی عملی توجه کنید: فرض می‌کنیم که دو تحقیق در یک‌سال درباره هشت شعبه از یک کتابخانه عمومی یا دانشگاهی انجام شده است (یکی در ژانویه و دیگری در ژوئن همان سال). گردآوری داده‌های مربوط به مدت زمان امانت (به روز) و تعداد دیرکردها (براساس درصد مواد به امانت رفته) بوده است. داده‌ها در جدول ۱ به نمایش گذاشته شده‌اند.

کتابدار براساس تجربه‌ها و مشاهدات شخصی خود احساس می‌کند که باید رابطه‌ای میان آن دو برقرار باشد، اما دلیل خاصی برای آن نمی‌یابد. در نگاه اول به جدول ۱ ممکن است بدون توجه به کوتاه یا بلند بودن مدت امانت به نظر برسد که کتابهای زیادی دیرکرد داشته‌اند، اما چیزی که مورد توجه کتابدار است، به‌ویژه در مورد دیرکردها، این است که چگونه مدت زمان امانت را به نحوی هماهنگ کند که هم برای استفاده کنندگان منصفانه و قابل تحمل و هم برای

جدول ۱. رابطه مدت امانت و تعداد دیرکردها (بررسی اول)

مدت امانت (به روز)	دیرکردها (درصد مواد به امانت رفته)	کتابخانه شعبه
۱۵	۳۸	۱
۱۸	۴۰	۲
۱۰	۵۶	۳
۷	۶۰	۴
۲۱	۷۵	۵
۱۳	۴۷	۶
۱۴	۵۲	۷
۱۲	۳۵	۸
(میانگین = ۱۳/۷۵)	(میانگین = ۵۰/۳۸)	

کتابخانه اقتصادی و قابل اجرا باشد (یعنی کتاب میان امانت‌گیرندگان بیشتری به گردش درآید). همان‌طور که در جدول ۱ مشاهده می‌شود راه یا دستورالعملی فوری وجود ندارد که مثلاً مدت امانت کوتاهتر دیرکرد کمتری را موجب شود. در حقیقت مدت امانت کتابخانه شماره ۷ دو برابر کتابخانه شماره ۴ بوده است، اما هر دو تقریباً به یک اندازه دیرکرد داشته‌اند. در حالی که کتابخانه شماره ۵ با طولانی‌ترین دوره امانت بالاترین میزان دیرکردها را داشته است.

برای تجزیه و تحلیل این آمار و ارقام راههای دقیق و صحیحی مورد نیاز است. در این مورد، محاسبه همبستگی یک شاخص معتبر است. براساس کتابهای آماری همبستگی انواع متفاوتی دارد. همبستگی اسپیرمن^۱ به صورت دستی به راحتی قابل محاسبه است، اما دقیق‌ترین همبستگی، همبستگی گشتاور پیرسون^۲ است که همه چیز را به صورت ریاضی محاسبه می‌کند. با استفاده از همبستگی پیرسون برای داده‌های این بررسی ضریب 0.153 به دست آمد. این ضریب همبستگی که به زبان ارقام بیان شده است نشان‌دهنده میزان رابطه میان متغیرهاست (در این بررسی رابطه میان دیرکردها و مدت امانت). ضرایب همبستگی چنانچه به درستی محاسبه شده باشند به صورت اعدادی بین $+1$ و -1 ظاهر می‌شوند که -1 ضریب همبستگی کاملاً منفی

(یعنی متغیرها در جهت مخالف یکدیگر حرکت می‌کنند) و $+1$ یک ضریب همبستگی کاملاً مثبت است (یعنی متغیرها در یک جهت حرکت می‌کنند، بالا یا پایین). هرگاه میان دو متغیر ضریب همبستگی صفر باشد به این معنی است که هیچ‌گونه همبستگی میان آن دو متغیر وجود ندارد.

همانند مقیاس ریشتر در زمین‌لرزه، مقیاس از $+1$ و -1 مقیاسی است که در آن، ارقام به سادگی بالاتر یا پایین‌تر نمی‌روند. در این مقیاس، ارقام یک حالت تصاعدی دارند، به طوری که مثلاً $0/2$ که نصف $0/4$ است خیلی از آن کمتر است، و $0/8$ خیلی بیشتر از دو برابر $0/4$ است. به این ترتیب به راحتی می‌توان گفت که مثلاً $0/6$ بسیار قویتر از $0/3$ است و غیره. این امر انسان را ترغیب می‌کند که ضرایب همبستگی را با نسبتهای قوت و ضعف مناسب تفسیر کند: مثلاً هر ضریب همبستگی بالاتر از $0/65$ ضریبی قوی (مثبت) در حالی که $0/27$ - ضریب همبستگی نسبتاً ضعیفی (منفی) است.

از محاسبه همبستگی‌ها می‌توان به رابطه میان اعداد هر ستون و میانگین آنها و نیز رابطه میان اعداد ستونها پی برده و واریانسها را تعیین کرد. باید این واقعیت را قبول داشت که این‌گونه محاسبات براساس مفهوم میانگین‌ها هستند و تنها به صورت شاخصهای ارزشمندی به حساب می‌آیند؛ زیرا به روش‌تر کردن مسئله کمک می‌کنند.

ضریب همبستگی $0/153$ نشان می‌دهد که رابطه‌ای میان دیرکردها و مدت امانت موجود است. این رابطه مثبت اما ضعیف است، یعنی تمایل ضعیفی وجود دارد که وقتی یکی از متغیرها بالا یا پایین می‌رود متغیر دیگر نیز بالا یا پایین برود. ولی از آنجایی که رابطه ضعیف است نمی‌توان چندان به آن تکیه کرد. همچنین باید قبول داشت که این تنها یک نمایش آماری از رابطه میان آن دو متغیر است.

در زندگی واقعی، ممکن است عوامل متعدد دیگری میان دو دسته از داده‌ها دخیل باشد که ما یقیناً از قبل آنها را می‌شناسیم: چیزی که نمی‌دانیم این است که آیا رابطه‌ای آماری میان آن دو دسته از داده‌ها وجود دارد یا خیر. محققان اغلب از داده‌ها به عنوان "متغیرها" یاد می‌کنند، یکی از متغیرها را به عنوان متغیر ناپیوسته (مثلاً فاصله محل کار یا محل سکونت استفاده کننده تا کتابخانه) و دیگری را متغیر پیوسته (مثل تعداد بازدیدها از کتابخانه) می‌نامند، پژوهشگران با بررسی رابطه میان متغیرها به ارزیابی خدمات کتابخانه‌ها می‌پردازند.

فرض کنید نتیجه تحقیق فعلی تصمیم‌گیری برای مدت زمان امانت باشد. ممکن است کتابخانه‌ها به دلایل گوناگون مانند: تأمین مدت زمان استفاده مناسب برای همه استفاده‌کنندگان، تنبیه یا تحقیق کردن، طول مدت زمان امانت را تغییر دهند. اطلاعات مربوط به بررسی دوم که

شش ماه بعد از بررسی اول انجام شده است در جدول ۲ منعکس است. از روی متوسط هر ستون می توان گفت که طول مدت امانت اندکی بیشتر شده است و میزان دیرکردها اندکی پایین آمده است. به هر حال متوسطها گویای ارزشهای واقعی نیستند و بسیاری از مردم ترجیح می دهند داده های واقعی را داشته باشند تا میانگین داده ها را.

جدول ۲. رابطه مدت امانت و تعداد دیرکردها (بررسی دوم)

مدت امانت (به روز)	دیرکردها (درصد مواد به امانت رفته)	کتابخانه شعبه
۱۶	۳۹	۱
۲۰	۲۵	۲
۱۲	۵۶	۳
۱۰	۵۸	۴
۲۵	۸۰	۵
۱۸	۴۰	۶
۱۷	۳۸	۷
۱۵	۲۷	۸
(میانگین = ۱۶/۶۳)	(میانگین = ۴۵/۳۸)	

مشاهده شرایط انفرادی هر کتابخانه، با مقایسه جدولهای ۱ و ۲، نشان می دهد که کتابخانه ۵، که با سخاوتمندی طول مدت زمان امانت را افزایش داده است، میزان دیرکردهایش نیز افزایش یافته است. استفاده کنندگان کتابخانه های ۷ و ۸ به نظر میرسد که مدت طولانی تر امانت را مفیدتر یافته اند. به منظور داشتن یک شاخص کلی می شود همبستگی را آزمایش کرد: همبستگی داده های این بررسی، ضریب $0/188$ را نشان می دهد. ضریبی که فقط اندکی قویتر از رابطه مثبت ضعیف، حاصل از بررسی اول (یعنی $0/153$) است. به عبارت دیگر، علی رغم رفتار انفرادی کتابخانه ها و علی رغم تغییرات طول مدت زمان امانت، هنوز این تمایل وجود دارد که دیرکردها و مدت زمان امانت با هم بالا یا پایین بروند.

برای درک بهتر مطلب لازم است که بیشتر به جنبه های عملی بپردازیم. در اینجا ما فقط یک شاخص آماری داریم. یقیناً اگر آن یک رابطه مثبت باشد ممکن است به این معنی باشد که

کاهش طول مدت زمان امانت می‌تواند به کاهش میزان دیرکردها بیانجامد. این یافته‌ای با ارزش است. ممکن است اختلافاتی میان شعب یا موارد مطالعه شده در ایام سال و میان گروههای استفاده‌کننده باشد، تمام اینها ارزش دارد که در تجزیه و تحلیل مورد توجه قرار گیرند. وجود رابطه مثبت ضعیف نشان می‌دهد که هرگاه مدت امانت طولانی‌تر شود، میزان دیرکردها نیز بیشتر می‌شود. به منظور اخذ اطلاعات گسترده‌تر در مورد این داده‌ها به تجزیه و تحلیل بیشتر می‌پردازیم.

تقسیم‌بندی دیرکردها براساس گروههای استفاده‌کننده

یکی دیگر از کاربردهای ساده آمار در جمع‌آوری داده‌ها براساس گروههای استفاده‌کننده است. استفاده‌کنندگان کتابخانه به هفت گروه سنی تقسیم شده‌اند: زیر ۱۰ سال، ۱۱-۲۰، ۲۱-۳۰، ۳۱-۴۰، ۴۱-۵۰، ۵۱-۶۰ سال، و ۶۱ سال به بالا.

برای این مطالعه، مواد به امانت رفته مربوط به متون داستانی بررسی و یکی از روزهای عادی هفته انتخاب می‌شود. البته شناسایی گروههای مناسب، رده‌های مختلف مجموعه کتابخانه، و زمان مناسب برای تحقیق به تصمیم‌گیری محقق بستگی دارد؛ و همچنین معیارهای متعدد دیگری از جمله عوامل تحقیقی و اداری نیز بر تصمیم‌گیری او تأثیر می‌گذارند. جدول ۳ نتایج این بررسی را نشان می‌دهد.

به منظور نمایش اهداف مختلف یک تحقیق، این بررسی خاص انتخاب شده است. این بررسی می‌تواند از طریق کنترل مواد امانت داده شده و برگشت آنها یا از طریق پرسش از استفاده‌کنندگان باشد داده‌ها می‌توانند داده‌های حقیقی باشند (مثلاً با استفاده از اطلاعات رفتاری، مثل

جدول ۳. بررسی میزان امانت داستانهای کوتاه

رده سنی	زیر ۱۰ سال	۲۰-۱۱	۳۰-۲۱	۴۰-۳۱	۵۰-۴۱	۶۰-۵۱	+۶۰
کتابخانه‌ها							
کتابخانه مادر	۲۳	۲۵	۴۳	۲۹	۱۹	۸	۳۲
شعبه ۱	۴	۱۰	۲۱	۱۳	۶	۱۰	۲۰
شعبه ۲	۲	۱۸	۲	۸	۱۱	۱۴	۱۲
شعبه ۳	۷	۳۲	۱۳	۱۸	۲	۷	۳۰
شعبه ۴	۱۰	۲۳	۱۴	۱۲	۲۵	۳	۲۴
شعبه ۵	۴	۴۲	۱۲	۸	۱۳	۱۴	۶

تعداد دفعات امانت گرفتن کتاب)، یا داده‌ها بر نظردهی باشد (مثلاً مردم درباره کتابخانه چگونه فکر می‌کنند). برای این‌گونه بررسی‌های اطلاعاتی، کدگذاری مناسب الزامی است تا داده‌ها بعد از گردآوری به صورت نظام‌یافته تجزیه شوند. در این بررسی خاص، رده‌ها شامل شعب یک کتابخانه و گروه‌های سنی استفاده‌کننده‌اند (جدول ۳). به منظور انجام هرگونه تحقیق به راحتی می‌توان داده‌های دیگری را جایگزین این رده‌ها (کتابخانه‌ها و گروه‌های سنی) کرد. مثلاً به جای شعب کتابخانه‌ها، رده‌های خاص داستانهای تخیلی (مثل پلیسی، عشقی، و غیره) یا به جای گروه‌های سنی، مواد اطلاعاتی به امانت گرفته شده (مشخص شده با شماره رده‌بندی دیویی) توسط این گروه‌های سنی، یا به امانت رفته از کتابخانه‌های مختلف را قرار داد. تقسیم‌بندی براساس جنس استفاده‌کنندگان نیز به‌سادگی میسر است.

بسیاری سؤالات مفید را کاتب‌دار می‌تواند براساس داده‌هایی همانند داده‌های جدول ۳ مطرح کند. مثلاً: شعبه ۱ در این بررسی چه وضعیتی دارد؟ مقایسه شعبه‌ها چه چیزی را نشان می‌دهد؟ وضعیت امانت برای افراد بالاتر از ۵۰ سال یا کودکان و افراد جوانتر چگونه است؟ با جمع ستونهای جدول ۳ که به‌سادگی میسر است می‌توان گفت چه مقدار از داستانهای تخیلی توسط افراد گروه سنی ۳۱-۴۰ در طول مدت معین به امانت گرفته شده است. با جمع‌بندی ستونها و ردیفهای جدول ۳، چه تعداد کتاب توسط کل گروه‌های سنی از هر یک از کتابخانه‌های شعبه به امانت رفته است، یا چند کتاب توسط کل کتابخانه‌های شعبه به گروه سنی خاصی به امانت داده شده است.

سؤال جالب دیگر این است که تا چه حد این داده‌ها انتظار ما را برآورده کرده‌اند. وقتی چنین سؤالی مطرح می‌کنیم، به‌طور ضمنی آنچه را اتفاق افتاده با آنچه را که انتظار داریم اتفاق بیفتد با هم مقایسه می‌کنیم. این اندیشه اساسی در آزمون میزان معنی‌دار بودن آماری است. در آمار چیزی معنی‌دار است، یا در سطح معینی معنی دارد، که زمانی اتفاق بیفتد که ما انتظار آنرا نداریم. در زندگی واقعی وقتی می‌گوییم چیزی "مهم" است که مسئله‌ای با اهمیت و غیرمنتظره اتفاق بیفتد. مانند وقتی که شخصی تغییر مهمی در وصیت‌نامه‌اش بدهد (مثلاً بر خلاف انتظار وارثانش، تمام اموالش را به یک بنگاه خیریه ببخشد) یا یک تغییر اساسی در وضعیت هوا اتفاق بیفتد، یعنی تغییری که به‌نظر می‌رسد تابستان تمام شده است.

از نظر آماری چیزی معنی‌دار است که در شرایط معینی در حالی که انتظار نمی‌رفته اتفاق افتاده باشد. وقتی می‌گوییم احتمال یک درصد دارد که این امر اتفاق بیفتد یعنی امر غیرمنتظره‌ای است، اما می‌تواند اتفاق بیفتد. به‌سهولت تجربی می‌دانیم که چه چیزهایی ممکن است به‌طور عادی اتفاق بیفتد، مثلاً نظم و به‌موقع آمدن اتوبوسها، قابل اطمینان بودن کار تهیه

کننده مواد کتابخانه، تعداد کتب به امانت رفته در یک صبح شنبه و نوع کتب مورد نیاز دانشجویان دوره لیسانس و غیره.

ما می‌توانیم از اینها به‌عنوان "داده‌های معلوم" یاد کنیم. این بسیار محتمل است که پراکندگی این نوع داده‌های معلوم شکل و فرم ویژه‌ای داشته باشند، شاید وضعیت عادی و غیرعادی را نشان بدهند. برای مثال ممکن است، ارقام نشریات به امانت رفته نشان دهد که بعضی اغلب به گردش در می‌آیند و پاره‌ای به ندرت به امانت می‌روند. یا "میانگین" مدارک به امانت رفته نشان دهد که مردم هر کدام ۴ کتاب به امانت برده‌اند، اگرچه بعضی هیچ کتابی به امانت نبرده‌اند و بعضی دیگر ۱۰ کتاب به امانت برده‌اند. این‌گونه دسته‌بندی داده‌ها را توزیع می‌نامند. مشهورترین توزیع آماری توزیع نرمال است، و اغلب منحنی آن در کتابها به‌شکل ناقوس ترسیم می‌شود. معمولاً منحنی بهره‌هوشی یک نمونه تصادفی بزرگ از کودکان، به همین شکل است. توزیعها انواع بسیاری دارند. برای مثال، هرگاه سکه‌ای را برای تعیین شیر یا خط به‌دفعات متعدد بیندازیم، یا با چشمهای بسته توپهای رنگی را از داخل ظرف پر از توپ رنگی برداریم، چیزی که به‌دست خواهیم آورد (مثلاً تعداد شیر و خط، یا ترکیبی از توپهای قرمز و آبی) تشکیل توزیع را می‌دهد. بعضی از ترکیبات کمتر از بعضی دیگر محتمل هستند، اما امکان تساوی هم وجود دارد: مثلاً احتمال اینکه ۵ توپ اول قرمز باشند یا ۵ سکه اول را که می‌اندازیم شیر باشند، از نظر آماری از احتمال مخلوطی از رنگهای قرمز و آبی و ترکیبات شیر و خط کمتر است.

اساساً توزیعها در زندگی واقعی یا در آمار، داده‌هایی را به‌ما می‌دهند که ما یافته‌های خود را با آن مقایسه می‌کنیم و همچنین درمی‌یابیم که آن چیزی که رخ داده است آیا قابل انتظار بوده یا غیرمنتظره است. فرق اصلی میان توزیعهای آماری و آنچه ما به‌صورت تجربی می‌دانیم، که گاهی توزیعهای تجربی نامیده می‌شوند، در این است که توزیعهای آماری براساس تصادف است؛ اما در هر دو حالت وقوع اتفاقات معمول و غیرمعمول جست‌وجو می‌شود.

در آمار هر چیز "در یک سطح معنی‌دار" دارای معنی است. سطوح مشهور معنی‌داری عبارتند از ۰/۰۵ (۵ درصد)، ۰/۰۱ (۱ درصد)، یا ۰/۰۰۱ (۱ در هزار)، تمام اینها تأکید بر عدم احتمال دارند، ما نیز بیشتر علاقه‌مند به وقوع اتفاقات غیرمنتظره هستیم تا منتظره! اگر گفتیم فرد "الف" که مسئول تهیه مواد کتابخانه است ۵ درصد کتابها را اشتباه می‌فرستد، می‌تواند خوب یا بد باشد. حداقل این فکر را به ما می‌دهد که تا چه حد کار او خوب بوده است. زیرا ما انتظار نداریم که اکثر کتابهای ارسالی توسط این شخص اشتباه باشد. دلایل ویژه‌ای ممکن است باعث ارسال کتابهای اشتباه شده باشد: برای مثال کسی سفارش را ترتیب داده باشد که به شدت در مورد شماره صورت‌حسابها بی‌دقت باشد.

منطق آمار به طور ساده این است که ما فقط انتظار استثناها را داریم، پس اگر بر خلاف انتظار ما عملی اتفاق بیفتد آن عمل معنی دار تصور می شود. بنابراین، چیزی معنی دار است که علی رغم غیرمعمول بودن وقوعش، اتفاق بیفتد (عملی که در سطح $0/01$ غیرمعمول است، در سطح $0/001$ به مراتب غیرمعمول تر است).

وقتی معنی دار بودن را آزمایش می کنیم، برای شروع کار به داده هایی احتیاج داریم که آنها را مشاهدات می نامیم و می توانیم آنها را محاسبه کنیم. اما در موقعیت های تصادفی آماری (مانند قبول نمونه معتبر)، ما علاقه مندیم که آنچه انتظارش را داریم رخ دهد. از راه های مختلفی می توان معنی داری را تعیین کرد. مثلاً در اینجا آزمون مربع کای انجام می گیرد. که با علامت یونانی (χ^2) آورده می شود.

بهترین حالت این است که داده ها در یک جدول توافقی (مانند جدول ۳) تنظیم شود. جمع ستونها و ردیفها نیز محاسبه شود. بعد زیر هر شماره، شماره دیگری ظاهر می گردد، شماره ای که تحت شرایط احتمال آماری است، داده ها را به صورتی که ما تصور می کنیم به ما می دهد (یعنی همان طور که آماردانان می توانند با یک "توزیع" نمایش دهند).

انحرافهای بین تمام این شمارهها محاسبه می گردد، و یک ارزش محاسبه شده برای مربع کای تعیین می شود. جدول ۴ آن را برای این بررسی برابر $129/325$ نشان می دهد. این شماره با جدول مقادیر بحرانی مربع کای (موجود در کتابهای آمار) مقایسه می گردد. نحوه مقایسه به این صورت است که: با تعیین یک سطح معنی دار (مثلاً $0/05$ یا $0/01$ و غیره)، از جدول مقادیر بحرانی مربع کای و انتخاب درجه آزادی^۱ قید شده در آن جدول، شماره مربع کای مربوطه در جدول را ثبت می کنیم (برای روشنتر شدن مطلب به این مثال توجه شود: طبق جدول مقادیر بحرانی مربع کای در کتابهای آمار، شماره های مربع کای در دو سطح $0/05$ و $0/01$ با 30 درجه آزادی به ترتیب $43/77$ و $50/89$ است). درجات آزادی کمک می کنند تا ویژگیهای خاص توزیع مورد استفاده را بارز کنیم. اکنون همه چیز آماده است تا رسیدگی شود که آیا بررسی از نظر آماری معنی دار است یا خیر.

قاعده این است که دو شماره (دو ارزش) مربع کای، یکی محاسبه شده از بررسی و دیگری موجود در جدول کتابهای آماری را با هم تطبیق داده و تعیین شود کدام شماره بزرگتر است. اگر شماره موجود در جدول کتابهای آماری بزرگتر بود، به این معنی است که یافته های بررسی معنی دار نیستند. یعنی ما باید انتظار داشته باشیم که یافته هایمان در شرایط احتمال آماری باشد.

جدول ۴. محاسبه مربع کای

گروه سنی	۱۰-	۱۱-۲۰	۲۱-۳۰	۳۱-۴۰	۴۱-۵۰	۵۱-۶۰	۶۱+	جمع
۱	۲۳	۲۵	۴۳	۲۹	۱۹	۸	۳۲	۱۷۹
	۱۳/۷۹	۴۱/۳۷	۲۰/۹۶	۲۴/۲۷	۲۰/۹۶	۱۵/۴۵	۳۴/۲۰	
۲	۴	۱۰	۲۱	۱۳	۶	۱۰	۲۰	۸۴
	۶/۴۷	۱۹/۴۱	۱۳/۵۹	۱۱/۳۹	۹/۸۴	۷/۲۵	۱۶/۰۵	
۳	۲	۱۸	۲	۸	۱۱	۱۴	۱۲	۶۷
	۵/۱۶	۱۵/۴۹	۱۰/۸۴	۹/۰۸	۷/۸۵	۵/۷۸	۱۲/۸۰	
۴	۷	۳۲	۱۳	۱۸	۲	۷	۳۰	۱۰۹
	۸/۴۰	۲۵/۱۹	۱۷/۶۳	۱۴/۷۸	۱۲/۷۶	۹/۴۱	۲۰/۸۳	
۵	۱۰	۲۳	۱۴	۱۲	۲۵	۳	۲۴	۱۱۱
	۸/۵۵	۲۵/۶۵	۱۷/۹۶	۱۵/۰۵	۱۳/۰۰	۹/۵۸	۲۱/۲۱	
۶	۴	۴۲	۱۲	۸	۱۳	۱۴	۶	۹۹
	۷/۶۳	۲۲/۸۸	۱۶/۰۲	۱۳/۴۲	۱۱/۵۹	۸/۵۴	۱۸/۹۲	
جمع	۵۰	۱۵۰	۱۰۵	۸۸	۷۶	۵۶	۱۲۴	۶۴۹

مربع کای =

$$\begin{aligned}
 & ۰/۹۴۴ + ۰/۱۴۲ + ۳/۵۸۹ + ۰/۱۸۴ + ۰/۹۲۱ + ۶/۸۰۷ + ۶/۴۷۸ + ۶/۱۵۰ \\
 & ۰/۴۰۸ + ۱/۹۳۷ + ۰/۹۳۷ + ۱/۰۴۵ + ۱/۴۹۶ + ۰/۲۲۸ + ۴/۰۴۰ + ۴/۵۶۵ + \\
 & ۱/۲۱۸ + ۱/۸۳۹ + ۰/۲۳۳ + ۰/۰۵۰ + ۱۱/۶۸۴ + ۱/۲۶۸ + ۰/۱۳۰ + ۷/۲۰۹ + \\
 & ۰/۶۱۸ + ۰/۸۷۳ + ۰/۲۷۵ + ۰/۲۴۵ + ۴/۰۴۱ + ۰/۶۱۵ + ۹/۰۷۸ + ۰/۷۰۲ + \\
 & ۰/۱۷۱ + ۲/۱۹۱ + ۱/۰۰۷ + ۱۵/۹۷۵ + ۱/۷۲۵ + ۰/۳۶۸ + ۴/۵۱۷ + ۱۱/۰۸۱ + \\
 & ۱۲۹/۳۲۵ = ۸/۸۱۸ + ۳/۴۸۷ +
 \end{aligned}$$

درجه آزادی = ۳۰

از طرف دیگر اگر مربع کای محاسبه شده بزرگتر باشد بیانگر آن است که علی‌رغم عدم درست‌نمایی یا احتمال (یعنی ۵ درصد)، این واقعه اتفاق افتاده است، و بنابراین، باید آنرا

معنی دار بدانیم. در این بررسی وضعیت این‌گونه است: ارزش مربع‌کای محاسبه شده $۱۲۹/۳۲۵$ است، در حالی که ارزش مربع‌کای جدول (در سطح $۰/۰۵$ معنی دار بودن و با درجه آزادی ۳۰) $۴۳/۷۷$ است. این به ما می‌گوید که داده‌های مورد استفاده معنی دارند، مخالف با آنچه ما باید از یک توزیع براساس شانس و تصادف انتظار داشته باشیم (نظیر یافته‌های برآورد شده از یک سلسله وسیع نمونه‌های تصادفی). حالا باید وضعیت را دوباره بررسی کرد تا دریابیم چه عاملی باعث این موضوع شده است. ممکن است گروه‌های دیگر را آزمایش کرد یا همین گروه‌ها را در زمان دیگر آزمایش نمود تا یافته‌ها تأیید شده و مطمئن‌تر و معتبرتر شوند.

وقتی که مردم در مورد قبول یا ردّ یک فرضیهٔ صفر^۱ صحبت می‌کنند وضع پیچیده‌تر است. این یقیناً مهم است وقتی که تصمیم می‌گیریم ببینیم یک فرضیه ویژه، صفر است یا خیر (مثلاً این فرضیه که افراد مسن بیشتر از کتابخانه کتاب به امانت می‌گیرند یا افراد جوان‌تر). مطالعهٔ دقیق کتابهای آمار برای درک چگونگی آزمایش فرضیه و میزان کردن آن با طرح تحقیق توصیه می‌شود. به‌علاوه، این مهم است که هنگام آزمایش کردن تمایل داشته باشیم که یک گروه از یافته‌ها بزرگتر یا کوچکتر از انتظارمان باشد. در آن صورت آزمون دو دنباله‌ای^۲ را انجام می‌دهیم. در صورت عدم پاسخ، امکان دارد استفاده از آزمون یک دنباله‌ای^۳ پاسخگو باشد. این اطلاعات در یک جدول خوب طراحی شده مربع‌کای موجود است.

شاید برای مردم عادی نسبتاً عجیب باشد که آماردانان از ردّ فرضیه صفر صحبت می‌کنند (یعنی فرض می‌کنند اختلافی وجود ندارد و سپس تلاش می‌کنند ثابت کنند که آن فرض نادرست است). اما آن داستان دیگری است و مهارت در این اندیشه‌ها در صورتی توصیه می‌شود که بخواهید در حوزهٔ تحقیقات تجربی تلاش کنید. حوزه‌ای که جزء تحقیقات عملی - پایه است و در کتابخانه و خدمات اطلاع‌رسانی کاربرد آن روند رو به رشدی دارد (طرح آنها از حوصله این مقاله خارج است).

ما باید در مورد یافته‌هایمان مراقب باشیم. آزمون خود به ما نمی‌گوید که چه چیزی باعث شده است که یافته‌ها معنی دار یا غیر معنی دار باشد. البته در کارهای متداول کتابخانه، چیزهای زیادی ممکن است باعث شوند که یافته‌ها معنی دار باشند. ممکن است شرایط ویژه‌ای تأثیر داشته باشد؛ استفاده کنندگان بخش داستانهای تخیلی ممکن است باهوشتر از معمول باشند؛ در میان آنها ممکن است نسبت (درصد) جوانان بیشتر از حد انتظار باشد؛ یا حتی تحقیق، خود موثق و معتبر نباشد (یعنی جهت دار باشد).

در چنین شرایطی، محقق بایستی به‌خوبی موقعیت را بررسی کند تا دریابد چرا و چگونه یافته‌ها معنی‌دار جلوه کرده‌اند، نتیجه تحقیق هرچه باشد، روشن است که تصمیم‌گیری به دنبال آن وجود دارد. اتخاذ هرگونه تصمیمی می‌تواند براساس تمام عوامل مربوط (ازجمله نتایج این تحقیق) و به ویژه تصمیمات مربوط به گروه‌های مختلف و خدمات ارائه شده به آنها باشد.

استفاده از مدرک و رگرسیون

یکی از وظایف دیرپای کتابدار این است که یقین حاصل نماید که کتاب مورد نیاز استفاده‌کننده به دستش برسد که مانند سایر بدیهیات اگرچه پیش پا افتاده است ولی اساسی است. در مدیریت مجموعه‌سازی یک مشکل اساسی این است که تنوع و توازن مجموعه در زمان و چارچوب برنامه‌ریزی شده فراهم شود. چگونگی استفاده از مجموعه نیز مهم است: آیا از آن استفاده می‌شود؟ به اندازه کافی استفاده می‌شود؟ چه قسمتهایی از آن بیشتر مورد استفاده واقع می‌شوند؟ چگونه مجموعه مورد استفاده واقع می‌شود؟ مردم انتظار دارند چه موضوعاتی را در مجموعه بیابند؟ و کتابدار از همه اینها چه استنباطی دارد؟

سؤال مشخص این است که چرا مردم از مجموعه کتابخانه استفاده می‌کنند یا نمی‌کنند. پاسخهای فراوانی وجود دارد: از فاصله استفاده‌کننده با کتابخانه تا توانایی خواندن او، مطالعه به‌منظور سرگرمی و یا تحت فشار یک دوره آموزشی، بی‌تفاوتی و یا تعهد خواننده، و ... از تصمیم موارد فوق سوالات ویژه‌ای مطرح می‌شود. یکی از این سوالات این است که یک کتاب جدید و با ظاهر مرتب و تمیز تا چه میزان به امانت می‌رود.

عمر مدرک می‌تواند بر میزان به امانت رفتنش تأثیر بگذارد: کودکان کتابهای حماسی را که مدت زیادی در کتابخانه بوده‌اند و کثیف هستند به امانت نمی‌گیرند؛ آنها کتابهای حماسی را ترجیح می‌دهند که مدت زیادی در کتابخانه بوده‌اند و وضعیت آنها نشان می‌دهد که توسط کودکان زیادی به امانت رفته و خوانده شده‌اند. برای استفاده‌کنندگان مسن‌تر، نو بودن فیزیکی و محتوایی کتاب ممکن است مهم باشد: بعد از همه اینها، هیچ‌کس نمی‌خواهد یک کتاب راهنمای قدیمی در رشته کامپیوتر را به امانت بگیرد. با استناد به مثالهای دیگر می‌توان یک بحث کامل از این موضوع را شامل انواع راههای مختلفی که یک مدرک کهنه می‌شود اعم از کهنگی فیزیکی یا محتوایی ارائه کرد.

این مسئله خاص را می‌توان به‌صورت آماری بیان کرد. فرض کنید ده مدرک در یک کتابخانه داریم، مثلاً آنها تک‌نگاشت چاپی در رشته زیست‌شناسی در یک دانشکده هستند. سن آنها را می‌توان بر مبنای تاریخ نشر یا تاریخ رسیدن و ثبت آنها در کتابخانه تعیین کرد. به‌منظور

جلوگیری از بحث و جدل شق دوم را انتخاب می‌کنیم. سپس فرض می‌کنیم که این کتابها برای یک دوره آموزشی مثلاً یک نیمسال تحصیلی مشخص شده‌اند و میزان به امانت رفتن آنها در جدول ۵ به نمایش گذاشته شده است.

جدول ۵. سن و میزان استفاده از کتابها

عمر کتاب (به سال)	میزان استفاده از کتاب (در سال)
۱	۵۹
۲	۶۵
۳	۱۷
۴	۲۸
۵	۲۵
۶	۱۶
۷	۱۵
۸	۱۰
۹	۳۱
۱۰	۱۴

این نمونه برای آمارگیری بسیار کوچک است. به‌طور مطلوب، نمونه‌ها باید بزرگتر باشند مثلاً بالای ۳۰ مورد. قاعدتاً نمونه بزرگتر نماینده بهتر جامعه‌ای است که از میان آن انتخاب شده است. به‌هرحال در اینجا برای این‌که روشها روشنتر نشان داده شوند، نمونه کوچکی انتخاب شده است. ما می‌توانیم از راههای مختلف این داده‌ها را تجزیه و تحلیل کنیم. مثلاً متوسط داده‌ها برابر با ۵/۵ برای سن کتاب و ۲۸ برای میزان استفاده: که می‌تواند شروعی برای تعمیم دادن رفتار و عمل متقابل داده‌ها بر همدیگر باشد. اما به‌منظور تجزیه و تحلیل بیشتر می‌توان همبستگی را آزمایش کرد. در این مورد ضریب همبستگی قوی منفی یعنی $-0/695$ - به‌دست می‌آید. این ضریب کمک می‌کند تا دریابیم که هر قدر که سن کتاب بالا می‌رود میزان استفاده از آن پایین می‌آید (یا هر قدر سن کتاب پایین باشد میزان استفاده از آن بیشتر است). انتظار می‌رود این همبستگی ادامه یابد مگر اینکه شرایط تغییر کند. گرچه برای اعتبار این بررسی لازم است که چندین نمونه دیگر مورد آزمایش قرار گیرند (البته بهتر است که نمونه‌ها تصادفی انتخاب شوند). به تجزیه و تحلیل بیشتر می‌پردازیم. با داشتن این داده‌ها، سودمند خواهد بود بدانیم که

انتظار داریم که یک کتاب چند بار به امانت برود (مثلاً هر کتاب ۶/۵ ساله). با آگاهی به اینکه نمونه کوچک است و ممکن است اعتبار نتیجه را محدود کند، رگرسیون می‌گیریم تا ببینیم چه می‌شود. رگرسیون از اساس احتمال آماری استفاده می‌کند. در کتابخانه به‌طور متداول، با شناختن رفتار قسمت‌های ویژه کتابخانه و گروه‌های استفاده‌کننده، به‌خوبی می‌توانیم نتایج را حدس بزنیم. اما با این داده‌ها و فرض اینکه در یک تحقیق واقعی داده‌های بیشتری خواهیم داشت، پاسخ را می‌توان به‌صورت آماری تعیین کرد.

رگرسیون بر این اساس کار می‌کند که هر جا دو متغیر باشد که بر یکدیگر تأثیر بگذارند و این تأثیر به‌صورت خطی دیده شود (یعنی X بیشتر، Y بیشتر، و غیره)، می‌توان با برآورد اطلاعات و مطالعه یکی از آن دو متغیر، به درک وضعیت متغیر دیگر پی برد. یک راه ساده درک مفهومی این موضوع با تفکر دربارهٔ دو متغیر مانند سن کتاب و میزان استفاده از کتاب است، و نیز تفکر در مورد چگونگی نمایش آنها روی یک منحنی ساده یا توزیع پراکندگی بر روی محورهای مختصات از طریق نقطه و خط‌هاست. با داشتن آن پراکندگی می‌توان دریافت که چگونه رابطه‌ای میان متغیرها وجود دارد و از آنجا تقریباً به‌صورت خام و ابتدایی دریابیم که از یک متغیر که درباره متغیر دیگر به‌ما اطلاعات می‌دهد چه انتظاری خواهیم داشت. رگرسیون این عمل را بسیار دقیق انجام می‌دهد و کاربردهای بسیار پیچیده‌تری را نیز می‌توان از طریق آن جست‌وجو کرد.

با کمک نرم‌افزاری مانند مینی‌تب می‌توانیم نتایج را به آسانی محاسبه کنیم. همان‌طوری‌که در دستورالعمل آمده است با گرفتن رگرسیون از متغیر دومی بر اساس متغیر اول، همواره پیشگویی متغیر دوم از روی متغیر اول ممکن است. بنابراین، براساس داده‌های خود مایلیم بدانیم که یک مدرک ۶/۵ ساله چند بار به امانت می‌رود. شمارهٔ مورد سؤال پیشگویی ۶/۵ است. پس از محاسبه، جواب، ۲۳/۶ یعنی برابر ۲۴ است، زیرا کسی نمی‌تواند کتاب را کمتر از یک‌بار به امانت بگیرد. کتابهای دستی و کتابهای راهنما مثالهای کاملی از روشهای محاسبهٔ رگرسیون را ارائه می‌دهند.

ممکن است محاسبه به‌صورت خلاف هم انجام شود. یعنی برای مثال مدرکی که ۲۲ بار به امانت رفته است چند ساله است. این بار ۲۲ را به‌عنوان اساس برای پیشگویی استفاده می‌کنیم. عمر کتاب برابر ۶/۱۶ یا تقریباً ۶ سال و دو ماه خواهد بود.

حتی ممکن است داده‌ها را برآورد (پیش‌بینی) کرد، اگرچه محدود بودن داده‌های این بخش از تجزیه و تحلیل را خیلی بی‌اعتبار می‌کند. برای مثال از یک مدرک ۱۵ ساله انتظار می‌رود که چند بار به امانت برود (جواب: تقریباً ۴ بار)، یا یک مدرک که ۷۵ بار به امانت رفته است چند

ساله است (جواب: تقریباً ۴ ماهه).

برآورد کردن با داده‌های محدود مانند این بررسی به زودی محدودیت‌هایش را نشان می‌دهد، به این صورت که شماره منفی می‌دهد (یعنی چیزی که در زندگی واقعی غیرممکن است). برآورد کردن داده‌ها روش مفیدی است، در صورتی که محدودیت‌هایش را بشناسیم. بعضی از این محدودیت‌ها به نوبه خود به داده‌هایی که در وهله اول جمع‌آوری شده برمی‌گردد و اینکه چه کاری قرار است با آنها انجام شود.

این روش ما را به یکی از بغرنج‌ترین جنبه‌های تصمیم‌گیری هدایت می‌کند، یعنی دانستن چیزهایی درباره آینده. اگر ما درباره رفتار داده‌ها واقع‌گرا باشیم و با مراقبت، آنها را در بستر اداره امور زندگی قرار دهیم، رگرسیون می‌تواند شاخصی مفید برای پیشگویی آینده باشد.

به ابتدای این مقاله برمی‌گردیم (مبحث بررسی دیرکردها و طول مدت امانت) تا رگرسیون را در آنجا آزمایش کنیم. در بررسی اول می‌توانیم چنین سؤالاتی را مطرح سازیم: وقتی که مدت امانت ۸ روز است چه درصدی دیرکرد انتظار داریم (جواب: ۴۷ درصد). در مقایسه با بررسی دوم همین سؤال را دوباره مطرح می‌کنیم، نتیجه دوباره ۴۷ درصد است. این نشان می‌دهد که تغییرات در آن حوزه از همبستگی تغییر واقعی را نشان نداده‌اند. همچنین انجام بسیاری از برآوردها ممکن است.

اگر از رگرسیون به صورت کامل استفاده شود روشی قدرتمند است. مخصوصاً وقتی ارزشمند می‌شود که کمیت مطلوبی از داده‌ها داشته باشیم و به خواهیم با استفاده از آن اطلاعات بیشتری در مورد آن داده‌ها به دست آوریم. ■

- در اینجا لازم می‌دانم از زحمات همکار محترم آقای زاهد بیگدلی که با دقت تمام متن را با ترجمه آن تطبیق داده و نکات مفیدی را یادآوری کرده‌اند صمیمانه تشکر کنم. - مترجم

برای مطالعه بیشتر:

Carpenter, R. L. ; Vasu, E. S. *Statistical Methods for Librarians*. Chicago: American Library Association, 1978.

Hafne, A. *Descriptive Statistical Techniques for Librarians*, Chicago: American Library Association, 1989.

Hennessy, L., *Quantitative Analysis*, Letts, London, 1989.

Monk, A., *Exploring Statistics with Minitab: A Workbook for the Behavioural Sciences*, Chichester: Wiley, 1991.

Owen, F. ; Jones, R. *Statistics*, 3rd ed. London, Pitman, 1990.

Phillips, J. L. *How to Think about statistics*, revised ed., freeman, Oxford, 1992.

Ryan, B.; Joiner, B; Ryan, T., *Minitab Handbook*, 3rd ed., Boston, pwskent Publishing MA, 1985.

Simpson, I. S., *Basic Statistics for Librarians*, 3rd ed., London, Bingley, 1988.

Simpson, I. S., *How to Interpret Statistical Data*, Library Association, 1990.