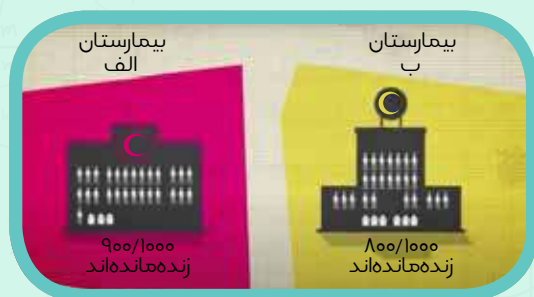




نقاط کور محاسبات از متغیرهای مخفی غافل نشویم

اگر حال بیمار وقتی به بیمارستان می‌رسد خوب باشد چطور؟ در بیمارستان ب ۹۸ درصد بیماران با وضع سلامتی نسبتاً خوب مداوا شده‌اند. با توجه به اینکه هم بیماران بدحال و هم بیماران سرپایی در بیمارستان ب بهتر مداوا می‌شوند، چطور بیمارستان الف می‌تواند نرخ بقای کلی بهتری داشته باشد؟ جواب این سؤال را باید در «ناسازنمای (پارادوکس) سیمپسون» پیدا کنیم!



علم آمار معمولاً قابل اعتماد است! خیلی از افراد، سازمان‌ها، و کشورهای تصمیم‌های بسیار مهمشان را بر پایه داده‌های طبقه‌بندی شده آماری می‌گیرند. ولی همین اطلاعات معمولاً قابل اتکا، گاهی به صورت پنهانی اشکال‌های بزرگی دارند! هر مجموعه آماری ممکن است خطاهایی ایجاد کند که مخاطب در نگاه اول متوجه آن‌ها نشود، ولی نتیجه را به کلی تغییر دهد. تصور کنید باید از بین دو بیمارستان «الف» و «ب» یکی را برای بستری کردن فردی از خویشاوندان خود انتخاب کنید.

از ۱۰۰۰ نفر بیماری که به تازگی به هر دو بیمارستان مراجعه کرده‌اند، ۹۰۰ نفر در بیمارستان الف به سلامت از بیمارستان خارج شده‌اند، در حالی که در بیمارستان ب ۸۰۰ نفر زنده مانده‌اند. به نظر می‌رسد بیمارستان الف انتخاب بهتری باشد! قبل از اینکه تصمیم نهایی را بگیرید، یادتان باشد همه بیماران با میزان سلامتی یکسان به دو بیمارستان نیامده‌اند.

اگر ۱۰۰۰ بیمار هر بیمارستان را به دو گروه تقسیم کنیم (آن‌هایی که با سلامت به نسبت خوب، و آن‌هایی که با سلامت کمتری به بیمارستان آمدند)، احتمالاً تصمیم شما متفاوت می‌شود. طبق اطلاعاتی که از دو بیمارستان گرفته‌ایم، بیمارستان الف تنها ۱۰۰ بیمار با وضع نامساعد داشته که ۳۰ نفر از آن‌ها زنده ماندند. اما بیمارستان ب ۴۰۰ بیمار با حال وخیم را پذیرش کرده است که ۲۱۰ نفر از آن‌ها نجات یافته‌اند. پس در بیمارستان ب ۵۲/۵ درصد بیماران بدحال مداوا شده‌اند که این عدد در بیمارستان الف حدود ۳۰ درصد است. پس برخلاف تصور اولیه، نرخ حیات در بیمارستان ب بالاتر است.



مثال دیگر، تحلیلی غلط در مورد حکم‌های محکومان در فلوریدا بود که ادعا شده بود هیچ اختلاف نژادی در صدور حکم بین متهمان سفیدپوست و سیاه‌پوست وجود ندارد. اما جداسازی پرونده‌ها بر اساس نژاد، این ادعا را رد کرد. چگونه می‌توانیم از افتادن در دام این تناقض‌ها جلوگیری کنیم؟ متأسفانه پاسخ خوبی برای این سؤال وجود ندارد! داده‌ها را می‌توان به روش‌های متفاوتی طبقه‌بندی کرد و از هر نوع طبقه‌بندی اطلاعات متفاوتی به دست می‌آید. گاهی ممکن است بررسی کل داده‌ها به صورت یکجا، تصویر و اطلاعات بهتری ارائه کند و گاهی تقسیم‌بندی داده‌ها به گروه‌های کوچک‌تر، اطلاعات دقیق‌تری در اختیار ما بگذارد. مهم آن است که به دقت مراقب متغیرهای پنهانی باشیم که ممکن است وجود داشته باشند. گاهی متغیرهای مخفی می‌توانند به کلی نتیجه و تحلیل اطلاعات را تغییر دهند. مراقب باشیم گول نخوریم!

گاهی استفاده از داده‌ها به تنهایی کافی نیست و لازم است چیزهای بیشتری از کل داستان بدانیم تا بتوانیم تصمیم درستی بگیریم. باید تصویر کامل و کافی از اطلاعات، مثلاً نحوه جمع‌آوری آن‌ها، عوامل مؤثر بر پاسخ و چیزهایی که بر هم تأثیر می‌گذارند در دسترس باشد تا موقع تصمیم‌گیری، چیزی از قلم نیفتد و اشتباهاتی مثل آنچه در مثال انتخاب بیمارستان دیدیم، رخ ندهد. در انتخاب بیمارستان، نسبت بیمارانی که با حال خوب یا حال بد به بیمارستان می‌رسند خیلی مهم بود که نزدیک بود ما به آن توجه نکنیم. ناسازنمای سیمپسون خیلی وقت‌ها در دنیای واقعی باعث بروز اشتباه‌های مشابه در تحلیل‌ها و تصمیم‌گیری‌ها می‌شود. تحقیقی نادرست در مدت ۲۰ سال در انگلستان انجام شده بود که طبق آن اعلام کردند افراد سیگاری بیشتر از غیرسیگاری‌ها شانس زنده ماندن دارند! نظر شما چیست؟ اشتباه تحقیق این بود که اطلاعات به‌دست‌آمده از تمام گروه‌های سنی را یکجا روی هم ریخته و آن‌ها را بررسی کرده بود. همان‌طور که می‌توانیم پیش‌بینی کنیم، بین غیرسیگاری‌ها، تعداد افراد بیشتری به سن پیری می‌رسند و ممکن است که در دوره انجام تحقیق فوت کنند. درست به دلیل اینکه آن‌ها زندگی طولانی‌تری نسبت به افراد سیگاری هم‌سن خودشان داشته‌اند و سیگاری‌های هم‌سن آن‌ها، قبل از رسیدن به سن پیری از دنیا رفته‌اند. در اینجا به گروه‌های سنی توجه نشده بود که این بی‌توجهی نتیجه را کاملاً تغییر داد!

