

نشریه علمی - ترویجی کارا



سال اول
شماره دوم
آذر ۱۳۹۷



صاحب امتیاز: انجمن علمی
آمار دانشگاه گیلان

معرفی نرم افزارهای آماری
آشنایی با داده کاوی و پروفسور نخعی زاده
معرفی دو کتاب آماری
آغازیک مسیر: اتحادیه دانشجویه آمار ایران

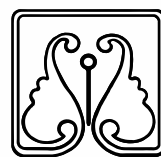
آیا رونالد فیشر راه می شناسید؟

روز جهانی آمار

مصاحبه اختصاصی با...

پارادوکس های اجتماع یافریبهای آماری؟

سواد آماری چیست؟



دانشگاه گیلان

دو فصلنامه علمی- ترویجی کارا

سال اوّل / شماره دوّم / مهر ۱۳۹۷
شمارگان: ۷۰ نسخه

مدیر مسئول: صادق بخشی پور
سردبیر: صادق بخشی پور*
صاحب امتیاز: انجمن علمی آمار دانشگاه گیلان
استاد مشاور انجمن علمی: دکتر رضا زارعی

هیئت تحریریه: (به ترتیب الفبا)

پروین اکبرپور

محمد اسمعیل زاد

صادق بخشی پور

لیلا حسن پور

علی حق پرست

سیده صدیقه ذکی پور

زهرا رافع

کسرا سهراباتی

امیر صالح غنجی

بنیامین لقمان نیا



همکاران این شماره:

الهه آزرمی و سیده محدثه حجتی

طراح جلد، طراح لوگو و صفحه آرایی:

ارسلان آریان مهر

عکاس: امیر صالح غنجی

آدرس پستی: رشت، بلوار نامجو، دانشکده علوم ریاضی دانشگاه گیلان، ساختمان کوشیار،

طبقه همکف، دفتر انجمن‌های علمی دانشکده علوم ریاضی.

صندوق پستی: ۱۹۱۴۱-۴۱۳۳۵

فهرست مطالب

سرآغاز

به نام خدای علیم و حکیم

شکر و سپاس و منت و عزت خدای را که جهان را چنین بر ستون های علم و حکمتش استوار نمود و سلام و درود بر محمد مصطفی (ص) همان ماحی، محوکننده ی ظلمت و آن هادی، معلم توانمند علم و بصیرت.

برخود میبایم و مفتخریم که بار دیگر فرصت خدمت به جامعه ی علم آمار را کسب نموده و توانستیم دومین شماره از نشریه علمی «کارا» را با رفع نقایص و کاستی های شماره اول و استفاده از انتقادات و پیشنهادات اساتید و دانشجویان منتشر کنیم. دنیای تقدیر و قدردانی را روانه ی دستان هنرمند و نکته بینی می کنیم که شماره قبلی این نشریه را متولد ساختند تا چراغی مبین باشد بر جاده ی علم و پژوهش در این رشته ی پرکاربرد و حیاتی. همچنین سر تعظیم فرود می آوریم بر شما دانشجویان با اراده که با پشت کار و عشق خود، ما را در این شماره همراهی نمودید.

بی تردید تمام توان خود را به کار بردیم تا رضایت شما عزیزان را کسب کنیم و امیدواریم نقص ها و کاستی های ما را با نگاه مثبت و بزرگوارانه خود پوشش داده و نظرات ارزشمند و سازنده خود را برای بهبود عملکرد نشریه در شماره های بعد به ما هدیه دهید.

به تحقیق اولین شرط موفقیت در رشته ای عشق به آن است اما این لازم کافی نیست و عشق بدون تخصص بی اثر و ناکارآمدست. سرآغاز تخصص آگاهیست که این آگاهی رسالت ماست.

همیشه یکی از مشکلات مهم آمار در کشورمان عدم توجه کافی به آن بوده است؛ علت آن هم درک کم برخی مسئولان از اهمیت و جایگاه آمار بوده است. به طوری که آمار برای آنها تنها وسیله ای بوده برای جبران کمبودها با اعلام آمارهای دروغ و در این میان کمتر نگاه علمی و درستی به آمار شده است؛ تنها در اداره آمار بانک مرکزی و مرکز آمار ایران این نگاه کارشناسی وجود دارد و در هیچ یک از دیگر ادارات و سازمانها چنین نگاهی به آمار وجود ندارد؛ در این ادارات کمتر دانشجوی آمار کاربرد استخدام می شود. اما بدون شک هر آنچه که ما امروزه آن را مشکل رشته آماروکاربردها می دانیم از بی عدالتی تا تبعیض، با اتحاد اعضای خانواده ی این رشته محو خواهد شد، بیایید این اتحاد را از همینجا -مرکز کسب علم و تجربه- دانشگاه تمرین کنیم. بنابراین بی صبرانه منتظر حضور شما در انجمن علمی آمار دانشگاه گیلان، از جمله نشریه «کارا» هستیم.

همچنین امیدواریم شاهد روزی باشیم که از «آمار» استفاده ابزاری نشود، کسی دروغ آماری نگوید و آمار درست مبنای تصمیم گیری مدیران کشور باشد. زیرا استفاده ابزاری از آمار و بزرگ نمایی یا کوچک نمایی با آن، ضربه ای اساسی بر پیکره فرهنگ آماری در کشور وارد کرده و سرمایه اجتماعی «اعتماد به گزارشات آماری» را از بین می برد.

در کلام آخر وظیفه خود می دانم از تلاش های بی وقفه هیأت محترم تحریریه پیگیری های مجدانه و دلسوزانه استاد مشاور انجمن علمی آمار دانشگاه گیلان جناب آقای دکتر رضا زارعی الطاف و کمک های اساتید گروه آمار دانشگاه گیلان؛ مخلصانه قدردانی کنم.

درپناه حق

صادق بخشی پور

مدیرمسئول و سردبیر نشریه علمی ترویجی کارا

پدر علم آمار جهان ۴

روز جهانی آمار و نگاهی به آمارهای بانکی ۶

آمار رسمی ۹

مصاحبه اختصاصی با دکتر صمیمی حق گذار ۱۲

پارادوکس سیمپسون ۱۵

پارادوکس های اجتماعی یا فریب های آماری ۱۷

سواد آماری ۲۲

لزوم سواد آماری در مدیریت ۲۶

آغاز یک مسیر: اتحادیه دانشجویی آمار ۲۸

معرفی نرم افزارهای آماری (۲) ۳۱

داده کاوی و آشنایی با پروفیسور نخعی زاده ۳۴

معرفی کتاب ۳۹

*در شماره اول نشریه «کارا» سرکار خانم طراوت نصیری سردبیر بوده اند اما در این شماره پس از اعمال تغییرات در تیم نشریه مدیرمسئول - جناب آقای صادق بخشی پور- به عنوان سردبیر این نشریه نیز تعیین گردیدند.

برای طرح انتقادات و پیشنهادات خود دربارهی نشریه و همچنین تمایل به همکاری با نشریه، با ما در ارتباط باشید.

Karajournal.guilanstat@gmail.com

ارتباط با مدیر مسئول و سردبیر نشریه:

Sadegh.bakhshipoor@gmail.com

وبسایت انجمن:

http://www.guilanstat.ir



پدر علم آمار جهان

زهرا رافع

تهیه کننده :



تقریباً هر روز در روزنامه ها و مجلات می خوانیم و یا در کتابهای علمی جملاتی را مشاهده می کنیم. مانند این که سیگار کشیدن موجب سرطان ریه می شود، ورزش کردن در جلوگیری از بیماریهای قلبی موثر است و غیره. دلیل درستی این اظهار نظرها چیست؟ شواهد آماری. برای تولید یک دارو شرکت داروسازی باید داروی خود را روی گروهی از بیماران آزمایش کند و در صورتی که نتایج آزمایشها نشان دهد مزایای دارو از نظر آماری معنی دار است اجازه عرضه آن را می دهند. اگر ذهن درخشان رونالد آیلمر فیشر نبود چنین نتایجی حاصل نمی شد.

رونالد فیشر دانشمندی است که علم آمار را پایه گذاری کرد، این نابغه تند مزاج کسی بود که نشان داد چگونه می توان از میان داده های خام حقایق محکم علمی را استخراج کرد. فیشر یکی از تاثیر گذارترین دانشمندان قرن بیستم بوده است. یافته های جدید بر پایه مفاهیمی استوارند که او ابداع کرده است مانند p -value یا تحلیل واریانس و ...

فیشر مردی که به علت بینایی بسیار ضعیف نمی توانست به میدان جنگ برود، با ماندن در کتابخانه و آزمایشگاه قوانینی را پی ریزی کرد که دنیای علوم تجربی را معنی دار کردند. برخی از یافته های فیشر آنچنان شگفت انگیزند که بیشتر به سحر و جادو می مانند تا ریاضیات. او روشی را یافت که از طریق آن می شد با داشتن نتایج سفرهای اکتشافی برای یافتن جانوران جدید، تعداد گونه های جانوری موجود در زمین را که هنوز کشف نشده اند، تخمین زد. او همچنین یکی از پیشگامان (نظریه مقدار کرانه ای) است. با استفاده از این نظریه، می توان نتایج رویدادهای گذشته - نظیر سیل ها یا زمین لرزه ها - را برای برآورد احتمال وقوع شدیدترین رویداد از این نوع طی هزار یا حتی ده هزار سال آینده، به کار گرفت. حتی تصور اینکه چنین پرسش هایی می توانند پاسخ داشته باشند، نیازمند خلاقیت و نبوغی سطح بالاست. اما فیشر این نبوغ را از کجا آورده بود؟ بدون شک پیشینه خانوادگی او در این مورد هیچ تاثیری نداشته است. فیشر در ۱۷ فوریه ۱۸۹۰، در فینچلی - شهرکی در لندن - به دنیا آمد. پدر او متصدی حراج اشیاء هنری بود و به جز یکی از عموهایش که در کمبریج ریاضیات خوانده بود، در خانواده آنها کسی به مقوله های علمی تمایل نشان نداده بود اما وی این مهارتها و علاقه مندی به اعداد را در سنین پائین از خود نشان داد. در مدرسه نیز استعداد او خیلی زود خود را نمایان کرد. اما در کنار آن، نقصی نیز عیان شد که ممکن بود پیشروی او در مسیر دانش را آهسته کند. اما برعکس، در فعالیت های حرفه ای او نقشی مؤثر و ارزشمند داشت. چشم های فیشر بسیار ضعیف بودند و در سنین نوجوانی این ضعف بینایی به حدی رسید که اگر از عینک های با عدسی بسیار ضخیم استفاده نمی کرد، تقریباً چیزی نمی دید. پزشک او نیز به او اخطار کرده بود که نباید زیر نور مصنوعی کار کند. این موضوع فیشر را مجبور کرد تا به جای استفاده از کاغذ و قلم، تنها از ذهن خود برای به تصویر کشیدن مسائل کمک بگیرد. در این میان، او روش های هندسی مخصوصی را ابداع کرد که به او امکان می داد مسائلی را حل کند که سایر ریاضی دانان که از روش های معمول استفاده می کردند، از حل آن ها درمانده بودند.

فیشر در سال ۱۹۰۴ وارد مدرسه هارو شد. به دلیل فوت مادرش در همان سال، برای او که یک پسر بچه ۱۴ ساله بود، دوران سختی به حساب می آمد. با وجود این، در سال ۱۹۰۶ در مسابقه ی مقاله نویسی مدرسه در زمینه ریاضی، برنده و موفق به دریافت مدال شد. به خاطر این موفقیت، به فیشر بورس تحصیلی دانشگاه کمبریج به مبلغ ۸۰ پوند داده شد. این مبلغ برای مخارج تحصیلی او درست زمانی که پدرش ورشکسته شده بود بسیار ضروری بود.

فیشر در حالی که در دو حوزه علمی که قرار بود نام او را بلندآوازه کند - ریاضیات و زیست شناسی - معلوماتی فراتر از هم سن و سالان خود داشت. اما زمانی که با رتبه اول در سال ۱۹۱۲ از این دانشگاه فارغ التحصیل شد، جایی را برای عرضه استعداد هایش نیافت دور از میدان جنگ در آزمایشگاه اندکی بعد با آغاز جنگ جهانی اول،

فیشر که علاقه مند بود به کشورش خدمت کند، برای رفتن به خط مقدم داوطلب شد. او همه آزمون ها را پشت سر گذاشت، اما از لحاظ میزان بینایی امتیاز لازم را کسب نکرد. به این ترتیب کسی که قرار بود برخی از مفاهیم کلیدی دانش در قرن بیستم را ابداع کند، از رفتن به جنگی که کشته شدن یا مجروح شدن در آن احتمال یک به سه داشت، معاف شد. به جای آن، فیشر به عنوان معلم مشغول به کار شد و در اوقات بیکاری نیز به بسط اندیشه های آماری اش پرداخت. سرانجام در سال ۱۹۱۹ شغلی را در یک ایستگاه تحقیقاتی در روتامستد در هرتفوردشایر - یکی از شهرستان های انگلیس - به دست آورد. در اینجا بود که اندیشه های آماری تحول برانگیزی که او مدت ها روی آنها کار کرده بود، کم کم شروع به نتیجه دهی کردند. در آن زمان، دانشمندان احساس می کردند که روش های آماری که در حال پدیدار شدن هستند، برای آزمون نظریات آنها مفید خواهند بود. با این حال، آنها نمی توانستند استفاده چندانی از این روش ها ببرند، چرا که داشتن حجم زیادی از داده ها برای بدست آوردن نتایج قابل احتمال. امری ضروری بود و دانشمندان چنین حجم عظیمی از داده ها را در اختیار نداشتند. فیشر تصمیم گرفت تا راههایی را برای استخراج قابل اعتمادترین نتایج از میان حجم کم داده ها بیابد. او دریافت که کلید این کار، بیرون کشیدن بیشترین حجم ممکن از اطلاعات از میان داده های موجود است. او زمانی که هنوز دانشجو بود، چیزی را یافته بود که به نظر میرسد دقیقاً برای این کار مناسب است. قاعده ریاضی «احتمال حداکثر»، چیزی بود که به فیشر امکان داد تا «آماره های کافی» را بیابد. اینها عبارت از کمیت هایی بودند که اطلاعات موجود در یک مجموعه داده را به بهترین شکل جمع بندی می کردند. فیشر که اکنون به ابزاری قدرتمند مجهز شده بود، روش های آماری متعددی را طراحی کرد که به دانشمندان آنچه را که مدتها در آرزوی بودند، عرضه می کرد؛ یافتن قابل اعتمادترین اطلاعات و نتایج از میان داده هایی که در اختیار داشتند. از کتاب علمی وی می توان روشهای آماری برای پژوهشگران را نام برد که در سال ۱۹۲۵ آن را به چاپ رسانید. این کتاب موهبتی الهی بود برای تحلیل داده ها و علی رغم انتقادها، روشهای وی به شکل گسترده ای مورد استفاده قرار گرفت و «آزمایش معنی دار بودن» برای کنترل نتیجه بر حسب واقعیت یا شانس بیشتر مورد تایید قرار گرفت.

فیشر متفکری تیز بین و فوق العاده بود. در عین حال نوشته های او برای بسیاری از خوانندگان قابل فهم نبود. بسیاری از ایده های فیشر به طور موثرتری، توسط کسانی که توانایی ساده سازی ایده های او را داشتند، به طور ساده تر بیان و در کتابها آمده است.

فیشر سخنان خوبی نیز نبود و درک سخنان او برای بسیاری از دانشجویان دشوار و غیر قابل فهم بود. معمولاً در کلاسهای او تنها ۲ یا ۳ دانشجو که توانایی درک وی را داشتند، به عنوان پیروان شیفته ی او باقی می ماندند. سرانجام در سال ۱۹۵۷ فیشر بازنشسته و به استرالیا مهاجرت کرد و در اثر ابتلا به سرطان از دنیا رفت. شاید کمتر کسی مانند او توان ریاضی خود را در اختیار همه انسانها گذاشته است.

بسیاری از واژه های آماری نظیر «تحلیل واریانس»، «آماره فرعی»، «سازگاری»، «کوواریانس»، «واریانس»، «کارایی»، «برآورد»، «درستنمایی»، «حداکثر درستنمایی»، «فرض صفر»، «توزیع نمونه ای»، «آماره»، «آماره ی بسنده»، «پارامتر»، «درجه آزادی»، «تواناترین آزمون»، «تصادفی کردن» نخستین بار توسط فیشر معرفی و به کار گرفته شد.

منابع و مراجع:

R. A. Fisher: The Life of a Scientist. Wiley. (۱۹۷۸) Box, Joan Fisher. ۱

۹۰۰۹۳۰۰-۴۷۱-۰۰ ISBN

A History of Mathematical Statistics. New York: (۱۹۹۸) Hald, Anders. ۲

۴-۱۷۹۱۲-۴۷۱-۰۰ Wiley. ISBN

blogfa.com.http://statistics۹۴.۳

روز جهانی آمارو
نگاهی به آمار های
بانك مرکزی

WORLD
STATIST
DAY
20.10.2015
BETTER DA
BETTER

پروین اکبر پور و لیلا حسن پور

تهیه کننده :



نام‌گذاری روزها با موضوعات خاص، سنتی است که هم در سطح ملی و هم در سطح بین‌المللی مرسوم است. روز جهانی آمار^۱ برای اولین بار در ۲۰ اکتبر ۲۰۱۰ در سراسر جهان جشن گرفته شد. این روز توسط کمیسیون آمار سازمان ملل متحد (UNSD) تعیین و نام‌گذاری شد. ر

بیش از ۱۰۳ کشور و حداقل ۴۰ سازمان منطقه‌ای و بین‌المللی روز جهانی آمار ۲۰۱۰ را جشن گرفتند^۲. به تصویب رساند که این رویداد هر ۵ سال یکبار برگزار گردد. دومین رویداد روز جهانی آمار روز ۲۰ اکتبر سال ۲۰۱۵، پنج سال پس از اولین روز جهانی آمار تحت تم « داده بهتر، زندگی بهتر» برگزار شد. هدف از برگزاری روز جهانی آمار، تاکید بر اهمیت استفاده از آمارهای رسمی در بین تصمیم‌گیران و سیاست‌گذاران در بالاترین سطح دولتی و توجه عمومی ملی به آن است. اول آبان هر سال در تقویم رسمی ایران نیز به روز ملی آمار و برنامه ریزی نام گذاری شده است.

* پیام دعوت دبیر کل سازمان ملل متحد از دولت‌ها برای برگزاری مناسبت روز جهانی آمار ۲۰۱۵

دبیر کل سازمان ملل متحد، آقای بان کی مون، در نامه‌ای به سران دولت‌ها در سراسر جهان در تاریخ ۱۹ آگوست ۲۰۱۵، از آنها دعوت کرد تا حمایت کامل و به موقع خود را برای برگزاری هر چه موفق‌تر روز جهانی آمار ابراز داشته و آنها را برای به اشتراک گذاردن وقایع این رویداد در جامعه جهانی، تشویق نمود. متن پیام:

نامه دبیر کل سازمان ملل متحد به دولت‌های سراسر دنیا

برای بزرگداشت روز جهانی آمار (۲۰ اکتبر ۲۰۱۵ میلادی و ۲۸ مهرماه ۱۳۹۴ شمسی)

عالیجناب،

افتخاردارم توجه شما را به بیانیه مجمع عمومی سازمان ملل متحد ۲۸۹/۶۹ که در آن مجمع تصمیم گرفت ۲۰ اکتبر سال ۲۰۱۵ را به عنوان روز جهانی آمار با موضوع «داده بهتر، زندگی بهتر» نام‌گذاری کند، جلب نمایم. مجمع عمومی سازمان ملل متحد از همه کشورهای عضو، مجامع تخصصی ملل متحد، سازمان‌های نظام ملل متحد و دیگر سازمان‌های بین‌المللی و منطقه‌ای و مجامع شهری شامل سازمان‌های غیر دولتی مانند مؤسسات تحقیقاتی، رسانه‌ها و تولیدکنندگان و کاربران آمارهای رسمی دعوت می‌کند روز جهانی آمار را گرامی دارند.

آمار نقش حیاتی در تصمیم‌سازی‌های شاهد- پایه در تمام زمینه‌های تاریخی و فرهنگی کشورها بدون وابستگی به سطح توسعه آنها ایفا می‌کند. با پذیرش قطعنامه ۶۹/۲۸۲، مجمع عمومی به اهمیت بنیادی ظرفیت آماری ملی ممکن در تولید آمارها و نشانگرهای قابل اعتماد و به موقع اذعان دارد. داده‌های ناب مبنای اغماض‌ناپذیر فرمولبندی سیاست‌های آگاهانه و نظارت بر توسعه‌های ۲۰۱۵ به بعد در سطوح ملی، منطقه‌ای و بین‌المللی هستند.

من از تلاشی که در سال‌های اخیر در بسیاری از کشورها در جهت ارتقاء ظرفیت آماری خود تحت رهبری دوایر آمار ملی، در زمینه‌هایی چون سرشماری نفوس و مسکن و نظارت بر رشد اهداف پیشرفت هزاره انجام شده است دلگرم شده‌ام. من همچنین از موفقیت قابل توجه بزرگداشت اولین روز جهانی آمار که

در ۲۰ اکتبر ۲۰۱۰ با فعالیت‌های سازمان داده شده در بیش از ۱۳۰ کشور عضو و حداقل ۴۰ سازمان منطقه‌ای و بین‌المللی برگزار گردید تشویق شده‌ام.

به منظور آنکه از روز جهانی آمار ۲۰۱۵ یک موفقیت مشابه بسازیم، از دولت شما به عنوان تولید کننده، کاربر و منافع شونده اولیه از آمارهای رسمی، دعوت به عمل می‌آورم تا پشتیبانی کامل و به موقع خود را در جهت مثمر ثمر واقع شدن تلاش‌های امسال شما نیز معطوف دارد.

بخش آمار دپارتمان اقتصاد و امور اجتماعی یک پایگاه اینترنتی برای به اشتراک گذاشتن اطلاعات مربوط به نحوه بزرگداشت روز جهانی آمار در سراسر جهان به آدرس <https://worldstatisticsday.org> ایجاد کرده است. مایلم دولت شما را تشویق کنم تا از این طریق ما را از برنامه‌های خود مطلع سازید تا آنها در معرض دید جامعه جهانی قرار گیرد.

بباید از این روز جهانی آمار تاریخی با اعلام و بزرگداشت نقش آمار در پیشرفت جوامع خود یک موفقیت بسازیم.

ارادتمند- بان کی مون

* نگاهی به آمارهای بانک مرکزی

همانطور که می‌دانیم بانک‌های مرکزی دنیا همواره در تلاش جهت ایجاد ثبات ارزش پول هستند که موجب ایجاد ثبات در سیستم مالی کشور توسط اجرای سیاست‌های پولی و اعتباری می‌شود. در این راستا نیازمند اطلاعات آماری شفاف، به هنگام، با قابلیت دسترس پذیری بالا و با کیفیت بوده و موظف شده اند با رعایت حقوق ارایه دهندگان آمار و محرمانگی، بطور مستقل و قدرتمند، نه تنها آمارهای دقیق مورد نیاز سیستم مالی و پولی کشور را تامین کنند بلکه به مسئولیت‌های اجتماعی خود نیز در اطلاع رسانی به عموم جامعه عمل نمایند. در این راستا تولید آمارهای مالی و پولی در دو حوزه آمارهای موسسات مالی-پولی و مالی-غیرپولی شاید مهمترین آمارهای تولیدی در بانک‌های مرکزی را در بر گیرند که شامل آمارهای مربوط به صورتهای مالی و نرخ سود این موسسات، آمار نرخ ارز و تبادلات و بدهی ارزی و شاخص‌های کلان اقتصادی برای سیاست‌های پولی، آمارهای اوراق بهادار و پورترفولیوها^۳ (portfolio)، آمارهای حسابهای مالی و صورتهای مالی تلفیقی موسسات غیرمالی برای سیاست‌های مالی، و همچنین تراز پرداخت ها، آمارهای بانکی بین‌المللی، بدهی خارجی و وضعیت سرمایه گذاری خارجی و غیره برای آمارهای ارزی و خارجی می‌شوند. بنابراین دور از انتظار نیست که بانک‌های مرکزی به یکی از مهمترین مراکز تولید آمار در همکاری صمیمانه با مراکز ملی آمار در کشورها تبدیل شوند و به عنوان منبع اصلی آمارهای مالی و بانکی محسوب گردند. این آمارها زیرمجموعه ای از آمارهای اقتصاد کلان کشورها است که معمولاً از مجموعه وسیعی از داده‌ها شامل فعالیت در بورس و سهام، گردش‌های مالی، نرخ‌های سود و شاخص‌های دیگر مربوط به فعالیت‌های مالی موسسات متعلق به بخش مالی اقتصاد استخراج می‌شوند تا بانک‌های مرکزی را قادر سازند پیشرفت و عملکرد کشور را در حوزه مالی کنترل و مشاهده نمایند و بر عملکرد بازارها و موسسات مالی اقدام نموده و آنها را در راستای افزایش کارایی و ثبات یاری رسانند که نظارت احتیاطی ناظر بر همین کاربرد است. اگر حوزه شرکت‌های مالی داخلی کشور را که مسئولیت تامین مالی و پولی را بر عهده دارند به دو حوزه موسسات مالی پولی در مقابل موسسات

^۳ پورترفولیو مجموعه ای است از اجزای متشکل از طرح ها، پروژه

ها، یا عملیاتی که برای رسیدن به اهداف استراتژیک به صورت

یک مجموعه واحد مدیریت می شوند.

^۱ World Statistics Day

^۲ United Nations Statistics Division

مالی غیرپولی در نظر بگیریم، مطمئناً بانک مرکزی به همراه بانکها و موسسات اعتباری و همچنین صندوقهای سرمایه گذاری بازار پول در دسته موسسات مالی پولی قرار می‌گیرند. اصطلاح «موسسه مالی پولی»^۴ بجای بانک و موسسات سپرده پذیر که در SNA ۹۳ (سیستم حساب های ملی)^۵ مطرح میشود طی اقدام اتحادیه اروپا در ایجاد حوزه پولی مشترک (پیمان ماستریخت)^۶ بکار گرفته شد و هدف این بود که این اصطلاح جدید، علاوه بر بانک مرکزی و بانکها، موسساتی مانند صندوقهای بازار پول یا صندوقهای مشترک سرمایه گذاری بازار پول را نیز در برگیرد. البته صندوقهای سرمایه گذاری بجز صندوقهای بازار پول، واسطه‌های مالی، شرکتهای بیمه و صندوقهای بازنشستگی هم در حوزه مالی فعال هستند که زیرمجموعه موسسات مالی غیرپولی قرار گرفته و اگر چه بطور مستقیم در حوزه پولی کشور حضور ندارند ولی در بازارهای مالی پولی کشور به سهم خود موثر هستند و توسط آمارهای بانک مرکزی رصد می‌شوند. یکی از اعضای موسس حوزه یورو در اروپا به نام الکساندر لامفالوسی^۷ در سال ۱۹۹۶ سخنرانی در رابطه با اهمیت آمار دارد و می‌گوید: «هیچ چیز به اندازه آمار خوب در سیاستگذاری پولی مهم نیست...» همچنین در سال ۲۰۰۸ آقای جین کلود تریچت^۸، رییس بانک مرکزی اروپا از سال ۲۰۰۳ تا ۲۰۱۱، می‌گوید: «اعتبار آمارهای حوزه یورو از مهمترین عوامل کلیدی مورد توجه بانک مرکزی اروپا است و این نشان می‌دهد که همانند گذشته می‌بایست اقدامات موثری در ابلاغ دستورالعملها و تقویت تولید آمارهای به روز صورت پذیرد تا بتوان بانک مرکزی اروپا را در اتخاذ سیاستهای پولی مناسب، بیش از پیش یاری رساند.» در همین راستا طبق مطالعه صورت گرفته قابل ذکر است که ۸۵٪ کشورهای حوزه یورو (بجز ایرلند، اسلونی و اسپانیا) و ۷۰٪ کشورهای اروپایی خارج از حوزه یورو و ۶۷٪ کشورهای دیگر عضو OECD^۹ دارای دپارتمان مستقل آمار با مدیریت متمرکز در بانک مرکزی خود هستند و بقیه کشورها دپارتمانهای مجزا در بخشهای مختلف مشابه بانک مرکزی جمهوری اسلامی ایران دارند. همچنین استفاده از فناوری روز دنیا و ایجاد انبارهای داده^{۱۰} مشترک و استاندارد و انجام تجزیه و تحلیل‌های پیشرفته مانند داده کاوی حجم زیاد داده^{۱۱} و ارائه اطلاعات توسط سامانه‌های یکپارچه هوش تجاری^{۱۲} به عنوان زیرساختهایی است که مورد استفاده بخش آمارهای بانکهای مرکزی دنیا قرار می‌گیرند. لازم به ذکر است که در ۲۰ اکتبر ۲۰۱۵ بانک مرکزی اروپا کنفرانسی را در فرانکفورت به مناسبت گرامیداشت سالروز جهانی آمار با شعار «داده بهتر، زندگی بهتر» با سخنرانی کلیه مدیران ارشد بانک مرکزی اروپا برگزار کرد که نشان دهنده اهمیت موضوع است. همانطور که مشخص شد حوزه‌های اعتبارسنجی، نظارت، سیاست‌های مالی و اقتصادی و ارزی بانکهای مرکزی، نیاز به تهیه آمارهای صحیح و به هنگام، جهت انجام ماموریت‌های خود دارند که نیازمند توجه به برخی نکات مانند موارد زیر است:

- وجود یک سیستم مدیریت اطلاعات جامع و یکپارچه مالی و پولی با استفاده از فناوریهای جدید
- بررسی و بومی سازی دستورالعملها و فناوریهای روز دنیا در رابطه با تولید، تبدیل و گزارش آمارهای مربوط به حوزه‌های اعتبارسنجی، نظارتی، ارزی، پولی و مالی و نظامهای پرداخت بانک.

- وجود همکاری صمیمانه با کل سیستم آماري کشور، در تولید آمارهای ملی و مورد نیاز و ایجاد چارچوبها و مفاهیم آماري مشترک ملی در حوزه مالی و پولی

- وجود قوانین و دستورالعملهای لازم جهت حمایت بانک مرکزی در تولید آمارهای مورد نیاز

- حفظ استقلال و حمایت از مراکز تولید آمارهای پولی و مالی بانک مرکزی در انجام ماموریت‌های محوله

- بکارگیری افرادی با تخصص بالا در حوزه آمارهای مالی و پولی در واحدهای مرتبط

- یکپارچه بودن آمارهای مورد نیاز بانک مرکزی از لحاظ ساختاری و فناوری جهت تهیه و تجمیع آمارهای حسابهای ملی، آمارهای اوراق بهادار، تراز پرداختها، موقعیت سرمایه گذاری خارجی، آمارهای مالی و پولی و ترازنامه ای و غیره.

منابع و مآخذ:

<https://www.linkedin.com>.۱

<https://worldstatisticsday.org>.۲

<http://src.ac.ir>.۳

<http://ecoview.ir>.۴

۴ MFI – Monetary and Financial Institutions

۵ سیستم حسابهای ملی از یک مجموعه منسجم، سازگار و یکپارچه حسابها، ترازنامهها و جداول مبتنی بر تعاریف و مفاهیم، طبقه بندی ها و قواعد حسابداری توافق شده در سطح بین المللی تشکیل یافته است.

۶ (Maastricht Treaty) پیمانی که در دهم دسامبر سال ۱۹۹۲

در شهر ماستریخت هلند توسط سران ۱۰ کشور عضو جامعه اقتصادی اروپا یعنی: آلمان، اسپانیا، انگلستان، ایتالیا، ایرلند، بلژیک، پرتغال، دانمارک، فرانسه، لوکزامبورگ، هلند و یونان به امضاء رسید.

۷ Alexander Lamfalussy

۸ Jean-Claude Trichet

۹ Organization for Economic Co-operation and Development

۱۰ Data Warehouse

۱۱ Big Data Mining

۱۲ Business Intelligence

آمار رسمی

علی حق پرست

تهیه کننده:



* مقدمه

ابتدایی، از هزاران سال پیش شکل گرفت. فرمانروایان برای نگهداشتن حساب خزانه، حسابدار داشتند که به ثبت دارایی های خزانه و تغییرات آن می پرداخت (آمارهای ثبتی)، برای آگاهی از توان نظامی ارتش خود به شمارش سربازان می پرداختند (سرشماری) و برای حفظ قدرت و گسترش آن به جمع آوری اطلاعات توسط خبرچینان اقدام می کردند (مونه گیری). به مرور زمان که جوامع گسترش یافتند و گروه های سیاسی، قومی، نژادی و ... شکل گرفتند، نیاز به برنامه ریزی برای هدایت امور جوامع و در نتیجه تقاضا برای کسب آمار و اطلاعات بیشتر احساس شد. شیوه های تولید آمار با گسترش و پیچیده تر شدن جوامع تکامل یافت و با توجه به گسترش دستگاه های حکومتی، زمینه های برنامه ریزی توسعه یافت و عرصه های جدیدی برای آمارگیری پدیدار گشت و به تدریج دستگاه های دولتی در قالب ضوابطی مأمور تهیه آمار واطلاعات شدند و در نتیجه نظام های آمار و نظام های تولید آمارهای رسمی همراه با نظام های برنامه ریزی تولد یافتند.

* معرفی برخی از درس های گرایش آمار رسمی

گرایش آمار رسمی تلفیقی از چندین مبحث آمار، اقتصاد، جمعیت شناسی، مدیریت و ... می باشد. در ابتدای ورود به رشته ی آمار رسمی، قبل از هر چیز باید شناخت مقدماتی از محیطی که در آن آمار های رسمی تولید می شوند پیدا کرد. در درس مقدمه ای بر آمارهای رسمی جایگاه اطلاعات آماری در برنامه ریزی ها و در پی آن به نقش نظام های آماری برای تولید و عرضه اطلاعات آماری و به تشریح وظایف واحد های مرکز آمار ایران و نقش هریک در تولید آمارها پرداخته می شود.

آمار های ثبتی بخش عمده ای از آمارهای رسمی را در دل خود دارند و نیز کاربردهای آمارهای رسمی در تحلیل های اجتماعی-اقتصادی مانند حساب های ملی معرفی می شوند. اما تولید آمار های رسمی به تنهایی کافی نیست و بررسی کیفیت این داده ها نقش مهمتری را ایفا می کند. به همین دلیل در رشته آمار رسمی، شناخت روش های مختلف آمار گیری، تفاوت آنها با یکدیگر، ابزار های گردآوری داده ها و مسائلی پیرامون آن بخش مهمی از مطالعات را به خود اختصاص می دهد که در درس روش های آمار گیری مورد بحث قرار می گیرد. یکی دیگر از مباحثی که کیفیت آمارهای رسمی را تحت تاثیر قرار می دهد، طرح های نمونه گیری می باشد. به همین جهت درس نمونه گیری کاربردی برای شناسایی انواع طرح های نمونه گیری، برآورد ها، نحوه برخورد با بی پاسخی و ... در نظر گرفته شده است.

بدیهی است آنچه که کیفیت داده ها را کاهش می دهد، بروز خطاست. در این رشته چگونگی جزئیات آمارگیری که بر کیفیت نتایج آنها اثر می گذارند و آنچه که در آمارگیری ها خطا شناخته می شود مورد بحث قرار می گیرد و به بررسی جنبه های مختلف طراحی و اجزا آمارگیری ها می پردازد. شناسایی انواع خطاهای آمارگیری اعم از خطاهای نمونه گیری و غیر نمونه گیری در درس خطای کل آمار گیری بررسی می شوند.

* هدف ها

هدف از تاسیس دوره کارشناسی ارشد آمار رسمی تربیت نیروی انسانی مورد نیاز نظام آماری کشور است. دانش آموختگان کارشناسی ارشد آمار برای جذب در دستگاه های تولید کننده ی آمار در نظام آماری ملی کشور باید دارای دانش و مهارت هایی باشند که در گرایش های فعلی کارشناسی ارشد آمار (آمار ریاضی و آمارجماعی-اقتصادی) به آنها توجه می شود. بنابراین نحوه تولید آمار به شیوه بسیار

آمار واژه ای است که معنایی دو گانه ولی مرتبط به هم دارد. در معنای اول آمار به «اطلاعات عددی» یعنی اعداد و ارقامی که به منظور خاصی تهیه و به شکل های مختلفی (موردار، جدول و...) نمایش داده میشوند، اطلاق میشود. در معنای دوم، واژه آمار، عنوانی برای یک شاخه ی علمی است که ناظر بر مجموعه فنون گردآوری، طبقه بندی و تلخیص داده ها و سرانجام استنباط آماری می باشد. ارتباط این دو معنا در این واقعیت نهفته است که اگر «اعداد و ارقام» پیش گفته با روش های علمی بدست آمده باشند باید با کمک علم آمار باشد. به عبارت دیگر، استفاده از علم آمار شرط لازم برای تهیه آمارهای قابل استناد است.

واژه ی آمار در عبارت «آمارهای رسمی» به معنای اول آن است ولی واژه ی رسمی، آن را به اطلاعات عددی خاصی محدود میکند. آمار های رسمی به اطلاعات عددی گفته می شوند که توسط دولت یا مراجع صلاحیت دار که در قوانین ومقررات مشخص هستند، تولید ومنتشر میشوند، و اطلاعاتی را در مورد وضعیت عمومی کشور برای امور مدیریتی (برنامه ریزی، سیاست گذاری وتصمیم گیری) بدست می دهند.

در تعریف واژه ی آمار به دو معنای آن و ارتباط آنها بایکدیگر اشاره شد. این ارتباط در نقشی نهفته است که علم آمار در تولید آمارها ایفا می کند زیرا بخشی از مطالعه های کاربردی این علم به حل مسائل واقعی و اجرایی در گردآوری داده ها وتولید آمارها اختصاص دارد. معمولا هدف این مطالعه ها ابداع روش هایی است که آمارهای با کیفیت بالا و باکمترین هزینه را بدست دهند. بنابراین باید در نظام آماری، تولید آمارها به متخصصان وآمارشناسان واگذار شود تا استفاده از روش های علمی در تولید آمار نمود پیدا کند. در این خصوص نیاز به رشته ای با نام «آمار رسمی» احساس می شود.

تولید آمارهای رسمی نتیجه فعالیت های مختلفی است: وضع قوانین ومقررات، تهیه ی تعریف ها و استاندارد ها، تعیین نیاز ها و اولویت ها، به کارگیری روش های علمی، طراحی ابعاد اجرایی، استخراج وتحلیل داده ها و سرانجام انتشار و اطلاع رسانی اطلاعات آماری.

مطالعه این فرایند آن چیزی است که در رشته آمار رسمی مورد بررسی قرار می گیرد. وظیفه مهم این رشته، پوشش این فرایند از طریق تأمین آموزش های تکمیلی به آمارشناسان است از آنجایی که فارغ التحصیلان گروه های آموزشی آمار اطلاعات زیادی درباره ی برآوردهای پیشرفته ی آماری دارند اما دانش کاربردی کمی پیرامون چگونگی طراحی نمونه گیری های ترکیبی، مواجهه با مشکلات پردازش داده ها، استفاده از آمار ثبتی و توانایی پیوند منابع آماری وچگونگی توسعه ی ابزارهای آمارگیری دارند، لازم است گرایش آمار رسمی بنیان گذاری شود که حوزه فعالیت آن تلفیقی از چندین مبحث آمار، اقتصاد، جمعیت شناختی، مدیریت، IT و ... می باشد. واضح است که ارائه حوزه های متنوع ذکر شده در قالب چند درس اختیاری که لزوما ارائه نمی شود، نمی گنجد و لازم است این مباحث در گرایشی مستقل پوشش داده شود.

* پیدایش آمار رسمی

از زمانی که جوامع تشکیل شدند و برای هدایت امور حکومت ها شکل گرفتند، حکمرانان برای اداره جامعه و گسترش اقتدار و گستره حکومت خود نیاز به برنامه داشتند و برای برنامه ریزی، خود را نیازمند شناخت وضع موجود یافتند که طبیعتا این شناخت با آمار حاصل می شد. بنابراین نحوه تولید آمار به شیوه بسیار

بتواند کمبود های فعلی و فاصله ای موجود در آموزش های تحصیلات تکمیلی آمار و نیاز نظام آماری ملی کشور را برطرف کند و همچنین بتواند آموزش های آکادمیک آمار رسمی را به کارکنان ارگان نظام آماری کشور ارائه کند هدف اصلی تاسیس گرایش آمار رسمی در مقطع کارشناسی ارشد آمار است.

* طول دوره و شکل نظام آموزشی

طول دوره ی تحصیلی در کارشناسی ارشد برای رشته ی آمار رسمی ۲ سال یا ۴ نیمسال در نظر گرفته شده است. طول عمر نیمسال ۱۶ هفته و هر واحد درسی به مدت ۱۶ ساعت در یک نیمسال خواهد بود. نظام آموزشی این رشته به صورت واحدی خواهد بود.

* ضرورت و اهمیت

یکی از وظایف نظام آموزشی عالی کشور تربیت نیروی انسانی کارآمد برای خدمت در بخش دولتی، بخش عمومی و بخش خصوصی است. دانش آموختگان نظام آموزشی عالی کشور باید دارای چنان دانش و مهارت هایی باشند که پس از جذب به دستگاه ها و سازمان های مرتبط با رشته ی خود علاوه بر آشنایی با مفاهیم و ادبیات مربوط قادر به همراهی با کارکنان با تجربه ی محیط کاری خود باشند و در شرایط ایده آل باید قادر به ایجاد تحول در نحوه ی انجام ماموریت های واحد خود باشند که به طور سنتی دنبال می شدند. دستگاه های نظام آماری ملی کشور سال هاست که در انتظار به خدمت گرفتن چنین نیروهایی هستند که هم با ادبیات آمار رسمی آشنا باشند و هم بتوانند با بکارگیری فن ها و مهارت های آموخته در دوره تحصیلات تکمیلی آمار رسمی شیوه های کاری سنتی نظام آماری ملی کشور را متحول کنند.

* اصول اساسی آمارهای رسمی

اصل ۱- ارتباط، بی طرفی و دسترسی برابر

آمارهای رسمی با ارائه اطلاعات در مورد وضعیت اقتصادی، جمعیتی، اجتماعی و زیست محیطی، جزئی ضروری از سیستم اطلاعاتی يك جامعه دموکراتیک را فراهم می سازد که هم دولت، هم اقتصاد و هم عموم افراد جامعه از آن بهره می برند. لذا قرار بر این است که آن دسته از آمارهای رسمی که سودمندی آن ها در عمل به اثبات رسیده است توسط مؤسسات آمارهای رسمی تهیه شده و بدون هیچ تبعیضی در اختیار علاقمندان قرار گیرد تا به این وسیله حق شهروندان در برخورداری از اطلاعات عمومی ادا شود.

اصل ۲- موازین و اخلاق حرفه ای

برای حفظ اعتماد عمومی به آمارهای رسمی، لازم است که مؤسسات آماری بر مبنای ملاحظات دقیق تخصصی شامل اصول علمی و اخلاق حرفه ای، در مورد روش های جمع آوری، پردازش، ذخیره سازی و توزیع داده های آماری تصمیم بگیرند.

اصل ۳- پاسخ گویی و شفافیت

برای تسهیل درک صحیح از داده ها لازم است مؤسسات آماری، اطلاعات را منطبق با استانداردهای علمی تدوین شده در مورد منابع، روش ها و روندهای آماری ارائه نمایند.

اصل ۴- پیشگیری از استفاده نادرست

مؤسسات آماری مجازند برای پیشگیری از تفسیرها و تعبیرهای اشتباه و استفاده نادرست از آمار، توضیحات لازم را ارائه نمایند.

اصل ۵- صرفه جویی

مؤسسات آماری می توانند داده های لازم را از انواع منابع، اعم از آمارگیری های مختلف و یا اطلاعات ثبتی اداری تهیه کنند. لازم است مؤسسات آماری منابع خود را با توجه به کیفیت، دقت زمانی، هزینه ها و زحماتی که متوجه پاسخگویان می شود، انتخاب نمایند.

اصل ۶- حفظ محرمانگی

اطلاعات مربوط به اشخاص اعم از حقیقی یا حقوقی که توسط مؤسسات آماری به منظور تهیه جداول آماری جمع آوری می شوند، باید کاملاً محرمانه محسوب شده و تنها برای مقاصد آماری مورد استفاده قرار گیرند.

اصل ۷- قانون آمار

لازم است قوانین، مقررات و موازین حاکم بر نظام آماری به اطلاع همگان رسانده شود.

اصل ۸- هماهنگی در سطح ملی

برقراری هماهنگی بین مؤسسات آماری يك کشور برای تحقق نظام آماری هماهنگ، منسجم و کارآمد، امری ضروری است.

اصل ۹- استانداردهای بین المللی

بهره گیری مؤسسات آماری کشورهای جهان از مفاهیم، طبقه بندی ها و روش های بین المللی، هماهنگی، انسجام و کارایی نظام های آماری را در تمام سطوح رسمی افزایش می دهد.

اصل ۱۰- همکاری بین المللی

همکاری دوجانبه و چندجانبه آماری موجب پیشرفت نظام آمار های رسمی کشور ها می شود.

این اصول دهگانه در نشست سال ۱۹۹۳ کمیسیون سازمان ملل تصویب شده است.

منابع و مراجع:

۱. سازمان برنامه و بودجه استان تهران (۱۳۷۵). مروری بر تاریخچه سرشماری ها و آمارگیری های ایران و جهان از دیرباز تاکنون.

۲. گزیده مطالب آماری، سال ۱۶، شماره ۴، زمستان ۱۳۸۴.

۳. How Are We Doing? Performance. (۱۹۹۷). De vries willem, F.M.

Indicators for National Statistical Systems. Netherlands Official statistics.

مصاحبه
اختصاصی با
دکتر صمیمی
حق گذار

امیر صالح غنجدی

تهیه کننده:



در شماره اول این نشریه مصاحبه ای داشتیم با دانشجویی پر تلاش و موفق جناب آقای سعید پورمند. اما در این شماره سعی کردیم گپ و گفتی اختصاصی و صمیمانه ای راجع به رشته آمار با جناب آقای دکتر صمیمی^۱ حق گذار - ریاست محترم دانشکده علوم ریاضی دانشگاه گیلان- به عنوان استادیار گروه آمار دانشگاه گیلان داشته باشیم. استادی بسیار خوش اخلاق، دلسوز و زحمتکش.

(روز ۲۰ ام مرداد ماه ۱۳۹۷)

**سلام آقای دکتر صمیمی، ریاست محترم دانشکده علوم ریاضی، خسته نباشد. روزتون به خیر.

- سلام به شما. بسیار ممنونم. شما هم خسته نباشید.

** لطفاً یه بیوگرافی از خودتون برای ما بگید.

- چی بگم اول؟ از شماره کفش پام شروع کنم؟ (با خنده). حضور شما عرض کنم که دیپلم بنده ریاضی فیزیک بود، لیسانس هم رشته ریاضی محض از دانشگاه گیلان سال ۱۳۶۵ و فوق لیسانس هم آمار نظری بود از دانشگاه شیراز سال ۱۳۶۹، بعد دیگه خدمت شما عرض کنم که PHD بنده هم نظریه احتمال بود در سال ۱۳۸۸. بعد از فوق لیسانس جذب دانشگاه گیلان شدم به عنوان هیئت علمی در سال ۱۳۷۰. از آن زمان هم تا به الان اینجا مشغول به خدمت هستم.

** آقای دکتر راجع به رشته آمار و کاربرد این رشته صحبت بفرمایید.

- آمار البته یک رشته ی نسبتاً نو پا و جدیدیه در ایران، همونطور که میدونید رشته آمار قبل از این که به صورت رشته معرفی شه، به صورت یک درس بود! درس آمار. که اونم اولین بار آقای دکتر خواجه نوری^۲ در دانشکده کشاورزی دانشگاه تهران که در کرج مستقر هست، در آنجا ایشان یک درسی گذاشت به نام آمار. حدوداً ۶۰-۷۰ سال قبل که از اونجا بود که آمار یواش یواش شکل گرفت. بعد اداره ثبت احوال به درس آمار علاقه نشون داد و درس آمار گذاشت و به همین ترتیب کم کم به صورت مستقل در دانشگاه های قدیمی تر مثل شهید بهشتی، شیراز و مشهد این رشته ارائه شد، یعنی درواقع این سه دانشگاه پیشگامان ارائه رشته آمار هستن. در مورد کاربرد این رشته حضور شما عرض کنم که همونطور خودتون میدونید این رشته کاربرد وسیعی داره در زمینه های مختلف، در کارهای تحقیقاتی و تقریباً همه ی رشته ها نیاز به آمار دارند. اینکه یک تحقیقات پزشکی قراره انجام شه، یک بخش آن تخصصیه و مربوط به خود رشته است ولی یه بخشی از اون باید از فیلتر آمار بگذره. برای مثال نمونه برداری میکنن، صحت و سقم کار رو بررسی میکنن. عمدتاً رشته آمار، اگر در یک جمله بخوام بگم کارش بررسی پدیده های نادقیقه، پدیده هایی که ما اصطلاحاً بهش میگیم تصادفی. پدیده های تصادفی زیادی داریم در عالم واقعی؛ در اقتصاد، در حوزه های مالی، بانکی، کشاورزی و حوزه های گوناگون و متنوع دیگه. به خاطر همینکه تقریباً همه ی رشته

ها آمار میخونن، حالا اندک رشته هایی مثل ادبیات فارسی ممکنه کمتر با آمار سر و کار داشته باشن ولی اکثر رشته ها آمار میخونن، یکسری رشته ها بیشتر و یکسری کمتر.

** چه شاخه ای از آمار مورد علاقه ی شماست؟

- حضور شما عرض کنم که بینید علاقه به مرور زمان تغییر میکنه و یه مقداری هم نگاه به شرایط بیرونی داره. در حال حاضر خیلی رغبت دارم به حوزه اقتصاد و مالی. اقتصاد اعم از بورس، بیمه و اینجور چیزها. بنابراین بیشتر الان متمرکز روی سری های زمانی مالی، یعنی مدل های سری های زمانی ای که به درد پیش بینی و مدل بندی پدیده های اقتصادی میخوره و در عین حال کاری حیاتی و دشواری هست. بعد مسائل ریسک در بیمه که اونم بی ارتباط نیست و تقریباً از یک خانواده هستن ولی رویکردشون و ابزار کاریشون اندکی با همدیگه تفاوت داره.

** راجع به تالیفات خودتون بگید آقای دکتر.

- در حال حاضر یک ترجمه در زمینه بیمه دارم که امیدوارم اگه تنبلی اجازه بده (با خنده)، البته که باید بگم اگه فرصت بشه ولی من رک میگم اگه تنبلی اجازه بده تا یک یا دو ماه دیگه سعی میکنم آماده اش کنم. چون الان رشته آمار تو این سرفصل های جدیدش یه درس به نام مبانی بیمه هم قرار داده شده به عنوان یه درس اختیاری و من احساس کردم کتابی به فارسی در این مورد وجود نداره از سال قبل کوشش کردم یه کتاب در این زمینه وارد بازار کنم که اونم امیدوارم یکی دو ماه آینده آماده شه. از تالیفاتم حضور شما عرض کنم که تا حالا ۳ تا کتاب منتشر شده، یکی مقدمه ای بر ترکیبات هست، یکی دیگه مبانی ترکیبیاته که به همراه آقای دکتر بی آزار^۳ نوشتیم و یکی هم مقدمه ای بر فرآیندهای تصادفی و کاربردهاست که اشتراکاً با آقای دکتر فتحی^۴ نوشتیم.

** دلیل عدم استفاده از تخصصان و فارغ التحصیلان رشته آمار در ادارات و سازمان ها، حتی در جاهایی که با آمار سر و کار دارند رو در چی میبینید آقای دکتر؟ چرا از افرادی با تخصص ها و با مدارک دیگری که هیچ ربطی با مقوله آمار، تحلیل و سواد آماری ندارند، استفاده می کنند؟

- ببینید ما سه جور برنامه ریزی داریم؛ برنامه ریزی سلیقه ای داریم، برنامه ریزی دستوری داریم و برنامه ریزی علمی داریم. برنامه ریزی علمی واقعا متکی به آمار هستش، یعنی اگر آمارهای واقعی به اصطلاح هم اخذ نشه و هم درصورتی که اخذ شده، مورد استفاده در برنامه ریزی ها قرار نگیره، اونوقت برنامه ریزی ها با چالش همراه هستن. علت استفاده نکردنش هم همینکه شاید این برنامه ریزان ما کاملاً لمس نکردن این مسئله رو که چقدر این موضوع حیاتی هست، یا حالا به هر دلیلی کم توجهی شده و الا اگر به این موضوع بیشتر توجه بشه و خوب بهش تکیه شه و درست استفاده کنن، ما تو برنامه ریزی ها کمتر دچار چالشهایی که الان میبینیم، نبودیم. همین! به نظر برنامه ریزان ما و سیاست گذاران ما اونطور که باید و شاید، به اهمیت

۱ samimi@guilan.ac.ir

۲ دکتر عباسقلی خواجه نوری (پدرآمار ایران) ، در ۴ تیرماه ۱۲۹۴ پا به عرصه وجود نهادند و در ۳۰ فروردین ۱۳۶۲ پس از سالها تدریس و تحقیق در سن ۶۸ سالگی بر اثر سکنه مغزی به رحمت ایزدی پیوستند.

۳ پروفیسور جعفر بی آزار، استاد دانشکده علوم ریاضی دانشگاه گیلان، گروه ریاضی کاربردی. (biazar@guilan.ac.ir)

۴ دکتر بهروز فتحی و اجارگاه، استاد دانشکده علوم ریاضی دانشگاه گیلان، گروه آمار. (fathi@guilan.ac.ir)

این رشته پی نبرده ان، حالا به کلام گفته میشه ولی در عمل چیز خاصی نمیبینیم.

****** آینده و سطح علمی آمار ایران رو چطور میبینید آقای دکتر؟

- آمار در ایران جایگاهش بد نیست به نسبت، یعنی هم متخصصان این رشته زیاد شدن، هم همکاران آماری ما حمله بردند به مرزهای دانش اصطلاحا و کارهای خوب و جدیدی انجام میدن ولی به نظرم یکمی فقط باید از حالت سنتی دربیاد، یعنی ما ها به عنوان یک آماری باید پیش از دیگران خودمون رو نقد کنیم و کوشش کنیم خودمون رو به روز تر کنیم و نسبت به مشکلات جامعه هم بی تفاوت نباشیم، یعنی باید راهکارهای علمی ما به سمت و سوی میل کنه که بتونه گره ای از مشکلات جامعه ما رو هم حل کنه و فقط کارهای نظری انجام ندیم، مثلا من مقدم به طور مشخص اینه که به مسائل تئوریک فقط در حد خودش بپردازیم و اینقد زیاد همش استنباط آماری، آمار ریاضی، یعنی به طوری که الان دانشجویهای کارشناسی آمار ریاضی میخونن، دانشجویهای کارشناسی ارشد، استنباط آماری میخونن که همونه، محتواش همونه، موضوعاتش حداقل همونه ولی به مقدار با عمق بیشتری. در مقطع دکتری آمار هم، استنباط آماری پیشرفته میخونن که اونم بازم همونه، ولی این باعث میشه که موضوعات جدیدتر و موضوعات دیگر مورد غفلت قرار بگیره. به مقدار برنامه ریزان و خود آماری ها هم به این توجه کنن که موضوعات و کارها رو یکمی متنوع کنن که بتونن به اصطلاح بتونن برای مشکلات روز هم راهکار داشته باشند.

****** آقای دکتر نصیحت شما به دانشجویان آمار چیه؟

- والا نصیحتی که دارم کلی هستش، فقط برای بچه های آمار نیست و برای همه دانشجویها هست. من تو کلاسام هم خیلی صحبت میکنم با دانشجویهام که دانشجو باید در درجه اول قدردان این موقعیتی که بهش داده شده باشه، یعنی اینکه بالاخره عمر هست دیگه، هر دانشجویی و هر کسی بهترین دوران عمرش رو داره در دانشگاه سپری میکنه و اگه از این موقعیت استفاده مفید نکنه، خسران محضه! باید از این وقت استفاده کنه، عمر دیگه برنمیگرده! خسران مالی قابل جبران، هزینه های دیگه قابل جبران اما عمر دیگه برنمیگرده و افراد باید سعی کنن از این وقتی که دارن میذارن، از این هزینه عمری ای که دارن انجام میدن، استفاده کنن و بیاموزن. این آموخته ها هدر نمیره. آموختن هیچ چیزی، هیچ چیزی (با تاکید) خسران نیست به نظر من. یعنی باید سعی کنن حتی الامکان از این عمری که گذاشتن بیاموزن. هر چی بیشتر، بهتر.

****** خیلی ممنون آقای دکتر از این وقتی که برای تیم نشریه کارا قرار دادین.

- تمام شد؟! من فک کرد سوالاتتون خیلی طولانیه (با خنده)

****** بله آقای دکتر موم شد، دست شما درد نکنه، میدونیم که خیلی سرتون شلوغه و بنابراین سعی کردیم سوالات مختصر باشه و بیشتر از این وقت شما رو نگیریم.

- ان شالله که همیشه موفق و موید باشد و با تلاش و کوشش بتونید مایه افتخار جامعه باشید.



پارادوکس سیمپسون

کسری سہراباتی

تمہیہ کنندہ :



پارادوکس سیمپسون: داده های زیادی همیشه مفید نیستند!

دانشگاهی میشود. اما با دانستن این مطلب و دیدن اطلاعات جدول ۲ چه نتیجه ای میگیریم؟ آیا میتوان از این آمار این نتیجه را گرفت که خواندن درس فیزیک دبیرستان، بطور میانگین باعث افزایش ۱۰٪ نمره درس فیزیک دانشگاهی، در میان دانشجویان تمامی رشته ها میشود؟ با دید معمولی این نتیجه گیری صحیح به نظر میرسد، اما کمی صبر کنید! بیاید این مسئله را به دقت بررسی کنیم. برای بررسی آمار این دو جدول، ما آنها را با هم در جدول ۳ و ۴ ترکیب میکنیم.

مجموع نمرات	میانگین نمرات	تعداد دانشجو	
4000	80	50	مهندسی
475	95	5	هنر
4475	-	55	مجموع
	81.4		میانگین

جدول 3.
میانگین نمره درس فیزیک برای تمامی دانشجویان با خواندن فیزیک دبیرستان

مجموع نمرات	میانگین نمرات	تعداد دانشجو	
350	70	50	مهندسی
4250	85	5	هنر
4600	-	55	مجموع
	83.6		میانگین

جدول 4.
میانگین نمره درس فیزیک برای تمامی دانشجویان بدون خواندن فیزیک دبیرستان

جدول ۳ آمار دانشجویان هر دو رشته را نشان میدهد که برای امتحان فیزیک خود درس فیزیک دبیرستان را نیز مطالعه کرده اند. همانطور که از این جدول پیداست در کل ۵۵ دانشجو اقدام به این کار کرده اند و معدل نمره این ۵۵ دانشجو ۸۱٫۴ شده است. اما جدول ۴، نشان دهنده ی آمار افرادی است که اقدام به خواندن درس فیزیک دبیرستان نکرده اند. بازهم تعداد ای دانشجویان ۵۵ نفر میباشد با این تفاوت که مجموع نمره این افراد ۸۳٫۶ شده است. مشاهده میشود که خواندن فیزیک دبیرستان در بین تمامی دانشجویان تأثیری در افزایش نمره نداشته است! اما در نگاه عادی ما به جداول ابتدایی، اینطور برداشت کردیم که خواندن فیزیک باعث افزایش ۱۰٪ نمره درس فیزیک دانشگاهی بین تمامی رشته ها میشود!

نتیجه گیری:

پارادوکس سیمپسون زمانی رخ میدهد که ما قصد داشته باشیم نتایج آماری مربوط به چند گروه مختلف با اندازه های متفاوت را بدست آوریم. مثال فوق به خوبی نشان میدهد که هنگامی که اندازه گروه ها متفاوت است نباید با یک دید کلی نتیجه گیری کرد، بلکه باید در این قبیل مسائل به دقت نتایج را بررسی کرد.

این موضوع که «هرچه تعداد داده های مورد بررسی از یک جامعه بیشتر باشد، نتیجه ی گرفته شده از آنها بیشتر قابل اطمینان خواهد بود» به صورت یک قانون قابل در آمد است. اما پارادوکس سیمپسون این قانون نانوخته را زیر سوال میرد! پارادوکس سیمپسون بر این نکته تأکید میکند که هنگامی که کار با داده های زیاد وجود دارد، مراقبت زیادی نیز میطلبد. در بعضی مواقع نتایجی که از داده های زیاد بوجود می آید، دقیقاً خلاف نتایجی است که از داده های کمتر بدست آمده اند. متأسفانه معمولاً نتایج بدست آمده از داده های زیاد، صحیح نیستند. برای بررسی این موضوع ما از یک سری داده های شبیه سازی شده استفاده میکنیم.

جدول ۱. اطلاعاتی است که از ۵۵ دانشجوی مهندسی بدست آمده است. این اطلاعات مربوط به نمره این ۵۵ دانشجوی مهندسی در درس فیزیک است. همانطور که پیداست ۵۰ نفر از این دانشجویان برای امتحان این درس، فیزیک دبیرستان را هم مطالعه کرده اند. میانگین نمره این ۵۰ دانشجو، ۸۰ (از ۱۰۰) شده است. همچنین ۵ نفر از این دانشجویان فیزیک دبیرستان را در کنار درس مورد نظر مطالعه نکرده اند که میانگین این دانشجویان ۷۰ شده است. بنابراین بر طبق این آمار، میتوان این نتیجه را گرفت که مطالعه کتاب فیزیک دبیرستان در کنار فیزیک (دانشگاهی)، نمره دانشجویان را به طور میانگین ۱۰ نمره افزایش داده است.

	با خواندن فیزیک دبیرستان	بدون خواندن فیزیک دبیرستان
تعداد دانشجویان	50	5
میانگین نمره	80	70

جدول 1.
میانگین نمره درس فیزیک دانشجویان مهندسی با / بدون خواندن فیزیک دبیرستان

در جدول ۲. میانگین نمرات ۵۵ دانشجوی رشته هنر در درس فیزیک آورده شده است. این درس برای این رشته به صورت اختیاری ارائه شده است. از میان این ۵۵ نفر، ۵ نفر برای شرکت در امتحان این درس، فیزیک دبیرستان را نیز مطالعه کرده اند. همانطور که پیداست در اینجا نیز نمره این ۵ نفر به طور میانگین ۱۰ نمره از نمره ۵۰ نفری که فیزیک دبیرستان را مطالعه نکرده اند بیشتر است.

	با خواندن فیزیک دبیرستان	بدون خواندن فیزیک دبیرستان
تعداد دانشجویان	5	50
میانگین نمره	95	85

جدول 2.
میانگین نمره درس فیزیک دانشجویان هنر با / بدون خواندن فیزیک دبیرستان

اگر اطلاعات جدول ۱ را داشته باشیم، میتوانیم این نتیجه را بگیریم که خواندن درس فیزیک دبیرستان، بطور میانگین باعث افزایش ۱۰٪ نمره درس فیزیک

پارادوکس های اجتماعی یا فریبهای آماري؟

امير صالح غنچي، محمد اسمعيل زاد

تهيه کننده :



* چکیده

آمار، یک دانش است. شعبده نیست. اما هر آماری بهر کاریست و بسته به نوع تهیه اش، شاخصی خواهد بود که نمی توان آن را در جای دیگر به کار برد. در این نوشتار، با اشاره به برخی ریزه کاریها، خواهیم دید که چطور با داده هایی یکسان (و فرضا درست) اما تحلیل های متفاوت، میتوان به نتایج آماری متفاوت (یا حتی ظاهرا متضاد) رسید. چطور میتوان زیر بیرق علم رفت و در عین حال، فریب داد و یا فریب خورد.

کلید واژه: پارادوکس سیمپسون - ویل راجرز - Gerrymandering - میانگین - پارادوکس اجتماعی - پارادوکس آماری

* مقدمه

دستکاریهای آماری سابقه ای طولانی دارند (به ویژه هر زمانی که انگیزه سیاسی نیز پشتوانه آنها شود) اما شاید سیمپسون (۱۹۵۱) نخستین کسی بود که تاثیر شیوه آمارگیری بر نتایج حاصل از آن (در جامعه ای واحد) را رسماً بررسی کرد و به روشنی نمایان کرد که چگونه میتوان با تقسیم داده ها به بخش های مجزا، به نتایج آماری مختلف و گاه متضادی رسید، چگونه میشود با نگرش فریبکارانه و یا صرفاً تبلیغاتی به آمار، برخی حقایق را در پس برخی دیگر، پنهان کرد.

نمونه هایی متنوع از اینگونه اختلافها را با یکدیگر بررسی خواهیم کرد. برای سادگی کار، جوامعی نسبتاً کوچک را انتخاب می کنیم اما نتایج نهایی به هر نوع جامعه ای قابل گسترش است.

*نمونه اول) چطور می شود میانگین درآمد را بیشتر از مقدار واقعی نشان داد؟

هر جامعه، مجموعه ای از افراد خانواده هاست و همینطور مجموعه ای از شهروندان. هر چند که این دو مجموع، به لحاظ تعداد با هم برابرند اما میتوانند نتایج آماری مختلفی حاصل کنند. جامعه ای را تصور کنید که سه خانواده A و B و C دارد. خانواده A، چهارنفره است و نسبتاً فقیر، خانواده B، دونفره و نسبتاً مرفه است. خانواده C، نیز سه عضو دارد و وضعیتی متوسط دارد. این جامعه در مجموع، ۹ شهروند دارد.

درآمد افراد خانواده A، {۲ و ۴ و ۵ و ۷} واحد است (واحد، میتواند «هزار تومان» یا «میلیون» و یا هر اندازه دیگر باشد). درآمد افراد خانواده مرفه B، {۲۰ و ۲۵} واحد و درآمد افراد خانواده C، نیز {۸ و ۱۲ و ۱۵} واحد است.

با این حساب، درآمد متوسط افراد این جامعه ۹ نفره، برابر خواهد بود با:

$$\frac{2 + 4 + 5 + 7 + 20 + 25 + 8 + 12 + 15}{9} \approx 10.9$$

اما برای کسی که بخواهد آن را بنا به دلایلی زیادتراً از مقدار فوق، جلوه دهد نیز راه حل های زیرکانه وجود دارد. مثلاً میتوان در عوض درآمد متوسط افراد، متوسط درآمد خانوادگی افراد را بیان کرد. درآمد متوسط هر فرد در خانواده اش، چنین است:

$$\begin{aligned} \bar{A} &= \frac{2 + 4 + 5 + 7}{4} = 4.5 \\ \bar{B} &= \frac{20 + 25}{2} = 22.5 \\ \bar{C} &= \frac{8 + 12 + 15}{3} \approx 11.7 \end{aligned}$$

پس میانگین درآمد خانوادگی افراد در این جامعه، برابر خواهد شد با:

$$\frac{4.5 + 22.5 + 11.7}{3} \approx 12.9 > 10.9 !!$$

عجیب نیست؟ مسلماً نه شعبده ای در کار بود و نه تناقضی. گذشته از اینها، حتی می توان متوسط درآمد خانواده ها را پیش کشید. این جامعه، سه خانواده در خود دارد پس میانگین درآمد خانوارها، برابر خواهد بود با:

$$\frac{(2 + 4 + 5 + 7) + (20 + 25) + (8 + 12 + 15)}{3} \approx 32.7$$

مسلماً همه این آمارها در جای خود صحیح اند. حيله آنجاست که یکی به جای دیگری مطرح شود و افسوس که این کار به سادگی شدنی است. چرا که تفاوت میان عباراتی مثل «متوسط درآمد افراد» و «متوسط درآمد خانوادگی افراد» و «متوسط درآمد خانوار» برای اکثر مردم، چندان روشن نیست و در ذهن اغلب افراد، اینها مفاهیمی مترادفند در حالی که چنین نیست.

پارادوکس های اجتماعی (Aggregation paradox) به اینجا ختم نمی شوند اما همه آنها در یک چیز، مشترکند که همان تاثیر تغییر مقیاس تحلیل (مثل تفاوت «خانه به خانه» یا «نفر به نفر» در مثال بالا) بر نتیجه نهایی است. در ادامه، نمونه گیری را مطرح و ضمن بررسی، به موضوع استقلال آمارها نیز خواهیم پرداخت. نشان می دهیم که آمارهای وابسته را نمیتوان به صورت جداگانه اعلام و از آن بهره جویی کرد.

*نمونه دوم) درآمد سرانه بیشتر اما تنگنای اقتصادی بیشتر در خانواده؟

آیا ممکن است که درآمد سرانه در یک کشور (نسبت به سال گذشته) بالاتر رود اما درآمد خانواده ها پایین تر بیاید؟ مگر میشود وضع مالی افراد بهتر شود اما در عین حال، خانواده ها بیشتر در تنگنا قرار بگیرند؟ در زیر نشان خواهیم داد که شدنیست.

برای سادگی، شهری را تصور کنید با ۱۵ خانواده شش نفره (جمعا ۹۰ نفر جمعیت) که در سال ۹۶، درآمد سرانه افراد این شهر، ۵ واحد بوده است. پس درآمد خانوارها در آن سال برابر بوده با ۳۰ واحد (۵*۶). به فرض اگر بیشینه قیمت اتوموبیل در این شهر، ۲۹ تومان باشد، فعلاً هر خانواده قادر است که یکی داشته باشد.

یکسال بعد (سال ۹۷)، درآمد سرانه بالاتر می رود و به ۷ واحد می رسد. ظاهراً به نظر می رسد که جامعه به سوی رفاه بیشتر پیش می رود اما درنگ کنید، هنوز برای نتیجه گیری قدری زود است. رفاه جامعه، به پارامترهای دیگری نیز وابسته است.

مسلماً در طول یکسال، جمعیت شهر پایسته نمی ماند. فرض کنید بر اثر پدیده های معمول اجتماعی (تولد / مرگ / ازدواج / طلاق / مهاجرت)، این شهر در سال ۹۷ به ۲۵ خانواده چهار نفره دگرگون شده باشد (جمعیت ۱۰۰ نفر). پس درآمد خانواده ها در سال جدید، برابر خواهد شد با ۲۸ واحد (۴*۷) و دیگر خانواده ها قادر نیستند همان اتوموبیل را بخرند. حتی اگر قیمتش ثابت مانده و تورم را نادیده انگاشته باشیم.

اکنون می توان این دوگانگی را به وضوح دید. هر چند درآمد سرانه

۸,۸ درصد بوده (۱۰۰* [۹۰۰۰۰/۸۰۰۰]). به همین ترتیب، جمعیت و تعداد بیکاران شهرهای B و C (در همان سال) طبق جدول پایین است:

آمار سال ۹۶			
	شهر A	شهر B	شهر C
جمعیت جوان	90000	100000	100000
جمعیت بی کار	8000	6000	4000
نرخ بی کاری در شهر	8.8%	6.0%	4.0%

در جدول بالا، نرخ بیکاری هر شهر، به صورت تفکیکی محاسبه شده. اما نرخ بیکاری کل کشور، برابر خواهد بود با:

$$\left(\frac{8000 + 6000 + 4000}{90000 + 100000 + 100000} \right) * 100 = 6.2\%$$

در بحث آمارهای وابسته، گفتیم که با گذشت زمان، پدیده های اجتماعی از قبیل تولد / مرگ / مهاجرت / ... بر جمعیتها تاثیر گذاشته و آنان را دگرگون می سازد. لذا فرض میکنیم در سال جدید (سال ۹۷)، جمعیت آن کشور، قدری افزایش و از ۲۹۰۰۰۰ نفر به ۳۰۰۰۰۰ رسیده باشد.

چنان که دیدیم، در سال ۹۶، شهر C به لحاظ نرخ بیکاری، وضع بهتری از دو شهر دیگر داشت. لذا انتظار می رود که در سال جدید، بخشی از جمعیت کشور، به خیال یافتن کار به آن شهر مهاجرت کنند. اکنون جمعیت و تعداد بیکاران در سال بعد (یعنی سال ۹۷) را نگاه می کنیم. تغییراتی که اشاره کردیم در آن مشهود است:

آمار سال ۹۷			
	شهر A	شهر B	شهر C
جمعیت جوان	50000	50000	200000
جمعیت بی کار	4500	3300	8800
نرخ بی کاری در شهر	9.0%	6.6%	4.4%

بیکاران از شهرهای خود به سوی شهر C سرازیر شدند. پس از قبل هم انتظار میرفت که نرخ بیکاری در شهر C بالا رود ولی جالب آنجاست که این مهاجرت، سبب نشد تا اوضاع دو شهر دیگر بهبود یابد. از مقایسه جداول، پیداست که نرخ بیکاری در هر سه شهر بالا رفت! و طبعاً نارضایتی نیز بیشتر! این شرایط، گاه به اثر ویل راجرز (will rogers phenomenon) نیز شناخته می شود. اما عجیب تر از همه، آن است که از دید دولت، اوضاع بهتر شده! بیاید نگاهی به نرخ بیکاری کل کشور در سال ۹۷ بیاندازیم:

$$\left(\frac{4500 + 3300 + 8800}{50000 + 50000 + 200000} \right) * 100 = 5.5\%$$

آری! نرخ بیکاری در کشور پایین آمده و از ۶,۲ درصد (در سال ۹۶)، به ۵,۵ درصد رسیده! هر چند که این یک حقیقت است اما دیدیم که حقایق دیگری هم در میان بود. میتوان تنها از کم شدن نرخ بیکاری کشور گفت و در عین حال، از افزایش بیکاری در شهرها، وضعیت مهاجرت و بسیاری حقایق دیگر، بی سر و صدا گذاشت و در یک کلام، تنها به گلچینی از آمارها، برای رسیدن به مقاصد تبلیغاتی، بسنده کرد.

افزایش یافت (از ۵ به ۷) اما درآمد خانواده ها کاهش یافت (از ۳۰ به ۲۸). قدرت خرید مردم بالا رفت اما قدرت خرید خانواده ها پایین آمد! [۱].

دیدیم که برخی از آمارها (مثل آمار درآمد و آمار جمعیت) چگونه به یکدیگر وابسته اند. جدا جدا نمی توان بر آنها دآوری کرد. نمیتوان به افزایش درآمد سرانه بالید اما در کنارش، از افزایش جمعیت گذشت. رفاه (یا هر پارامتر دیگر) در یک جامعه را نمیتوان تنها با استناد به یک آمار (و اشاره نکردن به آمارهای مرتبط) مقایسه و از آن برای تبلیغ، بهره جست. نمیتوان جلوی دوربین نشست و ده ها آمار پراکنده را (با چشم بستن بر ارتباطشان)، پشت هم ردیف کرده و رشد (یا افول) جداگانه هر یک را به رخ جهانیان کشید.

در مثال بعد، بیشتر به مساله تاثیر مقایس تحلیل داده ها پرداخته و توضیح خواهیم داد که هر کدام جایگاهی مجزا دارند که در آن، مفید خواهند بود اما اگر نابجا به کار روند، نمیتوانند دست آویزی برای فریبکاری قرار بگیرند.

***نمونه سوم** چطور می شود تعداد خانواد های تک نفره، بیشتر باشد و در عین حال، تعداد افراد متاهل، زیادتر؟! باز برای سادگی، شهری را تصور کنید که تنها ۲۵ خانه (و در نتیجه، ۲۵ خانواده) دارد. بطوریکه از این تعداد، ۱۰ خانه در اختیار ۱۰ زوج جوان و ۱۵ خانه باقی مانده در اختیار افراد مجرد است. اکنون می توان گفت که اکثر خانواده های این شهر، تک نفره اند (یعنی نسبت ۲۵/۱۵). اما در عین حال، اکثر افراد جامعه متاهلند. (نسبت ۳۵/۲۰)!

تناقض ظاهری نمایان شده در این مثال نیز همچون دو نمونه قبلی، از تقسیم اطلاعات آماری به بخش های مجزا سرچشمه می گیرد، اما کدامیک از این دو آمار درستند و کدام نادرست؟ مسلماً هر دو درستند اما هر یک در جای خود و به قصدی که گردآوری شده.

در یک جامعه واقعی، آمار نخست (۲۵/۱۵) میتواند بیشتر برای دولت یا بنگاههای معاملات ملکی اهمیت داشته باشد چرا که گویای این واقعیت است که بازار خرید و فروش مسکن، بطور عمده در دست قشر مجرد است و این در تنظیم قیمتها تاثیرگذار است. اما بالعکس، برای یک سازمان کمک های مدنی (که بیشتر در پی آمار تاهل و مجرد هستند)، مسلماً آمار دوم (۳۵/۲۰) سودمند خواهد بود. تمام این نتایج، از همان اصل نخستین که تغییر واحدهای اندازه گیریست (مثلاً از نفر به خانواده) سرچشمه می گیرد و تحت عنوان کلی «پارادوکس سیمپسون تعمیم یافته» شناخته و طبقه بندی میشود [۱] که در ادامه، نمونه های بیشتری از آن را خواهیم دید.

***نمونه چهارم** بیکاری در شهرها بیداد میکند. دولت اما میگوید نرخ بیکاری کم شده!

وضعیتی را بررسی خواهیم که در آن، نرخ بیکاری در تک تک شهرهای یک کشور، افزایش یابد اما نرخ بیکاری کل کشور کم شود!! کشوری نمادین را تصور کنید با ۲۹۰۰۰۰ نفر جمعیت جوان. این کشور، سه شهر دارد که آنها را با A و B و C نشان می دهیم.

گیریم که در سال ۹۶، جمعیت جوانان شهر A، ۹۰۰۰۰ نفر بوده که ۸۰۰۰ نفر از آنان بیکار بوده اند. پس نرخ بیکاری در این شهر، برابر با

این بار نیز آشکار شد که تکیه بر یک آمار (مثلا آمار بیکاری در کل کشور) و بسط ندادن آن به دیگر آمار مرتبط (از قبیل آمار تفکیکی / آمار مهاجرت / آمار رشد جمعیت / ...)، چگونه میتواند حقایق (یا بخشی از آن) را پنهان کرده و راه را برای سودجویی، هموارتر. تنها یک مثال بود. کشوری فرضی، آن هم تنها با سه شهر اما با اینهمه آشوب (یا ریزه کاری آماری). مشتی نمونه خروار. خودتان حدس بزنید که در کشوری بزرگ، چه داستانها که نمی توان سر هم کرد.

*نمونه پنجم) شهردار لایق!

شهردار دو شهر A و B، برای کسب عنوان «شهردار نمونه» با هم در رقابتند. در سال اول، شهردار A، ۱۰۰ طرح در دستور کار داشته که تنها موفق به انجام ۴۰ تا از آنها میشود (موفقیت ۴۰٪). در سوی دیگر، شهردار B (در همان سال)، ۵۰ طرح داشته که ۲۱ تا از آنها را به سرانجام رسانده (پیشرفت ۴۲٪). پس میتوان گفت که در سال نخست، شهردار B موفق تر عمل کرده:

$$100/40 < 50/21$$

در سال دوم، شهردار A، مجدداً ۵۰ پروژه تازه می گیرد و تنها ۵ تا را سامان می دهد (پیشرفت ۱۰٪) اما شهردار B، ۱۰۰ پروژه تازه گرفته که ۱۱ تا را پایان می دهد (پیشرفت ۱۱٪). پس در سال دوم نیز شهردار B موفق تر بوده:

$$50/5 < 100/11$$

حال، شهردار A که رقابت در هر دو سال پیایی را واگذار کرده، باید چه کند؟ او میتواند به جای عملکرد سالانه، عملکرد دو سال خود را پیش بکشد! اگر بررسی را بطور دوسالانه انجام دهیم، خواهیم دید که وضعیت وارونه میشود:

شهردار A، در طول دو سال، جمعاً ۱۵۰ پروژه داشته و جمعاً ۴۵ تا را پایان داده اما در مقابل، شهردار B، جمعاً ۳۲ پروژه از مجموع ۱۵۰ تا را انجام داده و داریم:

$$(100+50)/(5+40) > (50+100)/(21+11)$$

پس این بار، گویا اوضاع به نفع شهردار A چرخید! اما کدام یک از این بررسی ها درستند؟ مسلماً هر دو. و باز هم سرآغاز حيله آنجاست که یکی مطرح و دیگری پنهان شود. به لحاظ آماری، دید سالانه دارای ارزشی بالاتر است چرا که داده های آماری بیشتری در خود دارد و دید دوسالانه را نیز میتوان از آن نتیجه گرفت (اما چنانچه دید دو سالانه را داشته باشیم، نمی توان اظهار نظری بر دید سالانه داشت).

حال، شهردار لایق کدام است؟ پاسخ، تنها وابسته به این اعداد و نسبتها نیست بلکه بستگی به وعده های آنان هم دارد. شهردار لایق، آن است که وعده اش را عمل کرده باشد (خواه آن وعده، پیشرفت بیشتر در دو سال بوده باشد و خواه سالانه و یا هر دو). برای مثال، چنانچه شهردار A، از ابتدا وعده پیشرفت بیشتر در بازه دوسالانه را داده باشد، میتواند ادعای پیروزی کند اما در غیر آنصورت شانس نخواهد داشت. درباره شهردار B نیز همینطور است. هر ادعا می بایست تنها در بازه و همان شرایط وعده داده شده، بررسی شود که اگر چنین نباشد، همانگونه که در این نمونه دیدیم، هر کس ممکن است بتواند با تغییر بازه ها، خود را به نحوی شهردار لایق یا مستول دلسوز و یا دهها عنوان شبیه به اینها قلمداد کند.

در این نمونه، بودجه مالی دو شهرداری یکسان فرض شد تا (برخلاف دو نمونه پیشین) بتوان صرفاً به تاثیر تغییر مقیاس داده ها پرداخت. در غیر اینصورت، از هیچ یک از دو معیار فوق (سالانه و دوسالانه)، به تنهایی نمی توان برای داوری بهره گرفت و می بایست وارد بحث آمارهای وابسته نیز شد که پیشتر به آن اشاره و نمونه هایی آوردیم. جلوه ای جالب و دور از انتظار دیگری از این پارادوکس را می توان در [۴] یافت.

تا اینجا، در کنار معرفی آمارهای وابسته، مثال های مختلفی از تاثیر دسته بندی متفاوت داده ها (از فرد به خانوار. از شهر به کشور. از سالانه به دوسالانه) بر نتایج آماری را دیدیم. در آخرین نمونه، وجه دیگری از آن را نمایش می دهیم که به Gerrymandering شهرت



دارد.

*نمونه ششم) چطور یک اقلیت، میتواند بر اکثریت

اینجا نیز با تغییر در دسته بندی، خواهیم دید چگونه اقلیت یک کشور، می تواند بر اکثریت شهرها، اعمال نفوذ کند. کشوری را تصور کنید با ۱۵ نفر جمعیت که سه شهر دارد (هر یک، دارای ۵ نفر جمعیت). در این کشور، دو دسته A و B با یکدیگر در رقابتند. دسته اقلیت A را با دایره و دسته اکثریت B را با مربع نشان می دهیم که به اینصورت در کشور پراکنده شده اند:

اکنون مطابق تصویر زیر، سه محدوده (با جمعیت مساوی) روی آن در نظر بگیرید:

فعلا دیده میشود که در بیشتر محدوده ها، باز هم دسته B در اکثریت است. اما آیا همیشه همینطور است؟ بار دیگر همان کشور را (با همان پراکندگی افراد) بررسی کرده اما این بار در نواحی دیگر (جمعیت نواحی همچنان با هم یکسان است) :

اینبار، دسته A (با آنکه در اقلیت است)، بر اغلب نواحی مسلط است! بطور خلاصه، میتوان گفت که با ناحیه بندی اخیر، دسته B، اکثریت را در کشور و دسته A، اکثریت نفوذ در شهرها را پیدا کرده است.

اصطلاح Gerrymandering، برگرفته از نام البریج جری (Elbridge Gerry)، سیاستمدار آمریکایی و فرماندار ماساچوست است و مربوط به طرحی که وی با اعضای حزب متبوعش برای تغییر در نواحی رای گیری و به منظور افزایش نفوذ سیاسی ریخت. او را میتوان از قدیمی ترین پیشگامان سیمسون دانست. از قضا، یکی از آن نواحی، گویا قدری شبیه به یک سمندر بود و این دستمایه سردبیر روزنامه بوستون در سال ۱۸۱۲ گردید برای ابداع فعل to gerrymander که ترکیبی ابتکاریست از نام او (Gerry) و سمندر (salamander) [۲].

سخن آخر

از درآمد سرانه گرفته تا نرخ بیکاری، دیدیم که چیزها گاهی آنطور که به نظر می رسند، نیستند و عدم توجه به ریزه کاریهایی مانند «اثر تغییر مقیاس یا همان اثر سیمپسون» و «آمار وابسته»، چگونه میتواند حقایقی را پنهان یا وارونه جلوه دهد و در مجموع، چگونه میتوان از داده هایی درست، نتایجی اشتباه و یا ناقص گرفت. فریبکاری تنها دروغ گفتن نیست. نگفتن همه حقیقت و یا گفتن تنها بخشی از آن که به سود گروهی خاص باشد نیز گاهی فریبکاریست.

اگر بنا به تکیه بر آمار باشد (خواه برای هواداری و خواه برای انتقاد)، می بایست ظاهربینی و نگاه احساسی کنار رود. انتقاد چشم بسته همانقدر آسیب رسان است که طرفداری کورکورانه.

آمار، همچنان حیطة خوبی برای ماهیگیران آبهای گل آلود است اما از سوی دیگر، نقش ساده باوران و خوش خیالان که خواسته یا ناخواسته، رونق بخش بازار فریبکاران هستند را نیز نمی توان نادیده انگاشت. با اندک بهره ای از ریاضی میتوان رویارویی اغلب چنین فریبکاری هالیستاد اما در واقع، همواره هستند افرادی که آسان فریب می خورند، زیرا که دوست دارند فریب بخورند.

منابع و مراجع:

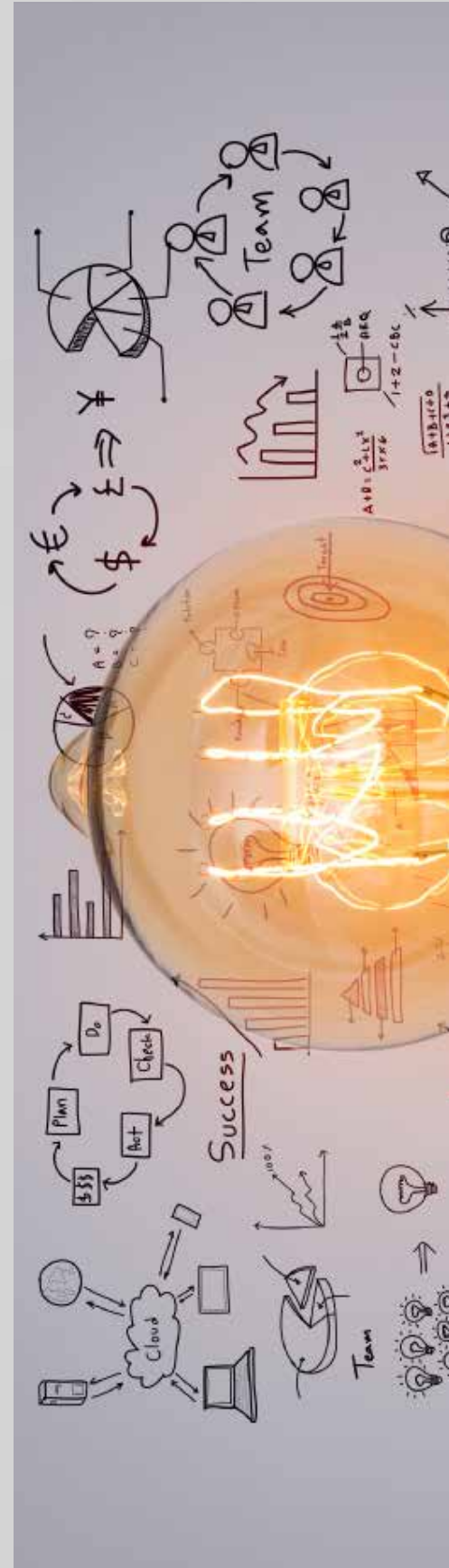
[۱] Collective Decision Making: Views from Social Choice and Game Theory

[۲] The rise and development of the gerrymander

[۳] The sum of the parts can violate the whole. American Political Science Review

[۴] Good for Women, Good for Men, Bad for people. Journal of Women's Health &

Gender Based Medicine



سواد آماری

تهیه کننده :

صادق بخشی پور



و پایه ای برای تصمیم سازی درمورد سیاست عمومی مثل تعیین محدوده های انتخاباتی و تعیین مکان مدارس و بیمارستان ها و ... ایجاد می کند. به کمک این اطلاعات آماری، صاحبان کسب و کار نیز می توانند بازارهایشان را بشناسند، کسب و کار و استراتژی های بازاریابی خود را گسترش بخشند.

در جامعه ی سرشار از اطلاعات امروز، داشتن سواد آماری، یک مزیت محسوب می شود و شما را برای کارفرمایان آتی فردی جذاب تر می کند و در محل کارتان بالاتر از سایر رقبا قرار می دهد. با افزایش دانش آماری خود، شما می توانید با اقتدار، دقت و درستی وارد مباحث و مراحل تصمیم گیری شوید.

* آیا شما سواد آماری دارید؟

تنها شما نیستید که استفاده از آمار برایتان راحت نیست. بسیاری از مردم به علت پیچیده بودن آمار، از آن فراری هستند. ممکن است افراد ندانند کجا به دنبال اطلاعات مورد نیاز خود بروند، با مفاهیم آماری آشنا نباشند و نسبت به توانایی خود در درک اعداد اعتماد نداشته باشند. برای کار کردن با آمار نیازی نیست حتما یک متخصص آمار یا ریاضی باشید. ریاضیات و کار با اعداد به معنی داشتن توانایی اولیه ی کار با ریاضی و درک مقدماتی از اعداد و ارقام است و مسلما پیش نیاز سواد آماری است. منظور از سواد آماری، ماهر بودن در فرمول بندی یا درک روش شناسی موجود در پشت سر اعداد نیست، بلکه منظور توانایی تفسیر اعداد و برقراری ارتباط صحیح بین اطلاعات موجود در آن است. آمار، به بیان ساده ی یک داستان کمک می کند. این آمارها میتوانند به روش های مختلفی مثل جداول، نمودار ها، نقشه ها یا متون ارائه شوند و اگر مفهوم آن ها را بدانیم به هیچ وجه ترسناک یا خسته کننده نیستند.

استفاده زیاد از آمار لزوما به افزایش درک ما از آن منجر نمی شود. در این دوران مملو از اطلاعات، مهم است که هر فردی دارای استقلال و تفکر انتقادی باشد و برای تحقق آن، داشتن سواد آماری ضروری است. اما برای استفاده از آمار باید توجه داشت که غفلت از تعاریف و مفاهیم یا مقایسه ی آمارها به طور نادرست، منجر به تفسیر نادرست داده ها خواهد شد.

* معیارهای داشتن سواد آماری

برای باسواد بودن از لحاظ آماری، نیاز به کسب مهارت در چهار حوزه ی مهم دارید:

آگاهی داشتن نسبت به داده ها

آیا داده ها، مرتبط و مناسب هستند؟ داده ها پایه و اساس آمار میباشند. آنها مشاهداتی هستند که پس از سازمان دهی و ارزیابی، تبدیل به اطلاعات و دانش می شوند. تفسیر داده ها نیاز به یک رویکرد سیستمی دارد. ابتدا درمورد سوالاتی که برای پاسخ به آن نیاز به داده دارید فکر کرده و موارد زیر را در نظر بگیرید:

• آیا داده ها از یک منبع معتبر هستند؟

• آیا داده ها واقعا معرف مطالب مورد نظر می باشند؟

• چرا داده ها به این شیوه ارائه شده اند؟ و به چه داده های دیگری نیاز است تا پاسخ کاملی به سوالات داده شود؟

یک نکته مهم از جواب سواد آماری، درک این موضوع است که چه چیزی داده ها را قابل اعتماد می سازد. درک چگونگی تولید داده ها، ما را از قضاوت های آگاهانه درمورد کیفیت داده ها مطمئن می کند.

داده ها از کجا می آیند؟ داده ها ممکن است از منابع متعددی گرفته شوند. از موارد زیر دوری کنید:

داده های موجود از قبل: این داده ها ممکن است برای یک هدف خاص تولید شده باشند. ممکن است جمعیتی که داده ها مبتنی بر آن بوده با جمعیت فعلی متفاوت باشد یا ممکن است شیوه ی نمونه گیری برای مطالعه ی فعلی مناسب نباشد.

داده های ثانویه: از این داده ها ممکن است در جایی استفاده شود که برای

منظور از سواد آماری چگونگی استفاده از آمارها است. از این رو می توان سواد آماری را به عنوان موضوعی در علم روش ها نیز تلقی کرد زیرا «چگونه فکر کردن» و «چگونه استدلال کردن» موضوع علم روش هاست. پایه و اساس مفاهیم سواد آماری بیشتر مبتنی بر استدلال استنتاجی است. در استدلال استنتاجی، اعتبار یک استدلال را بین صفر تا یک ارزیابی می کنیم. در حالی که در استدلال استقرایی یک استدلال یا درست است یا غلط. داستان طنز شکارچی که در ادامه به آنها اشاره می شود، تفاوت استدلال های استنتاجی و استقرایی را در ارتباط با سواد آماری روشن می کند.

* داستان طنز شکارچی:

دو شکارچی هنگام گردش در جنگل متوجه شدند که یک خرس گرسنه آنها را دنبال می کند. شکارچی اول به دلیل نداشتن سواد آماری طعمه خرس شد و شکارچی دوم به دلیل داشتن سواد آماری توانست از مهلکه جان سالم به در ببرد.

شکارچی اول به استناد درستی این آمار که سرعت دویدن خرس دو برابر سرعت دویدن یک انسان است از دویدن ناامید شد و طعمه خرس شد (استدلال استقرایی). شکارچی دوم با این استدلال که علیرغم اینکه سرعت دویدن خرس دو برابر سرعت دویدن یک انسان است ولی برای اینکه بتوانم خودم را نجات دهم، نیازی نیست که سرعت دویدن من بیشتر از سرعت دویدن خرس باشد، فقط کافی است سرعت دویدنم کمی بیشتر از سرعت دویدن شکارچی اول باشد (استدلال استنتاجی) نجات یافت.

سواد آماری بر تصمیم گیریها با استفاده از آمار به عنوان سند و مدرک متمرکز شده است، همانگونه که سوادخواندن و نوشتن بر استفاده از کلمات متمرکز شده است. سواد آماری به ما کمک می کند تا داده های زندگی خود را به اطلاعاتی قابل اعتماد و درست تبدیل کنیم و توسط آن برنامه ریزی صحیحی در زندگی داشته باشیم

با سواد آماری کسی است که علاوه بر استفاده صحیح از اطلاعات پیرامون خود بتواند اطلاعات صحیح را از اطلاعات نادرست تشخیص دهد. با سواد آماری کسی است که با ترفندهای جمع آوری و تجزیه و تحلیل اطلاعات آماری آشنایی داشته و به راحتی تحت تاثیر اطلاعات نادرست قرار نگیرد. اهمیت این مطلب تا آنجا است که در حال حاضر در کشورهای توسعه یافته داشتن سواد آماری همتر از سواد خواندن و نوشتن تلقی شده و علم آمار را در رده بیست تحول اول زندگی بشر در هزاره قبل قرار داده است. سوال اینجاست که سواد آماری ما تا چه اندازه است؟

* در بحث زیر مفصل به این پرسش خواهیم پرداخت.

از نظر اچ. جی. ولز^۱، روزی فرا خواهد رسید که تفکر آماری همچون خواندن و نوشتن ضرورت می یابد. و الآن آن روز فرا رسیده است. ما روزانه در ارتباط با ارقام و اطلاعات هستیم. بسیاری از عناوین خبری، به طور منظم گزارشات آماری شامل مطالب زیر هستند:

۱. تخلفات رانندگی طی پنج سال اخیر ۲۵ درصد افزایش یافته است.

۲. از هر پنج ازدواج، یکی منجر به طلاق می شود.

۳.

آمارها داستان های جالبی به ما می گویند که به کمک آنها می توانیم جهان را بهتر درک کنیم. آنها نشان دهنده ی تغییرات هستند. از طریق آن می توان مقایسه های مهم و قابل فهمی انجام داد. برای آن که قضاوت های درستی داشته باشیم لازم است از بهترین و دقیق ترین دانش برای اهداف پژوهشی و برنامه ریزی و تصمیم گیری برخوردار شویم.

* چرا داشتن سواد آماری ضروری است؟

تهیه ی اطلاعات آماری درست و دقیق و معتبر، جامعه ی ما را قدرتمند می سازد

هدف یک گزارش یا مطالعه ی خاص متناسب باشد. در این حالت ممکن است منبع داده ها قابل اعتماد نباشد. به عنوان یک اصل کلی، در صورت امکان از منبع اصلی داده ها استفاده کنید.

داده های تولید شده از مشاهده یا آزمایش: نوع سوالات مورد پرسش و نحوه ی پرسیدن سوالات می تواند جواب های دریافتی را تحت تاثیر قرار دهد. داده ها را می توان از کل یک جمعیت یا از یک نمونه که می توان نتایج آن را به جمعیت بیش تری تعمیم داد، جمع آوری کرد. انواع نمونه گیری ها می تواند متغیر باشد اما باید به خاطر سپرد که هر نمونه باید نماینده ی آن جمعیت بوده تا منجر به نتایج قابل اعتماد و معتبر شود.

شواهد داستان گونه: منظور رویداد خاص است که به طور معمول معرف کل نیست اگر چه برای توصیف یک نمونه ی خاص مفید است، اما باید به هنگام نتیجه گیری در مورد جامعه ی بزرگتر دقت کرد.

داده های اریبی: اریب یا تمایل به یک طرف، به طور عمدی یا سهوی در برخی آمارگیری ها وجود دارد. منابع اریب عبارتند از: اریب نمونه (آیا اندازه ی نمونه مناسب است و چگونه پاسخگویان جامعه انتخاب می شوند؟)

خطاهای پاسخ (افراد ممکن است سوالات را نادرست تفسیر کنند و جواب های صحیح ندهند)

داده های از قلم افتاده (افراد ممکن است به همه ی سوالات پاسخ ندهند یا اطلاعات ناقص بدهند)

پاسخ ها ممکن است تحت تاثیر جمله بندی سوالات قرار گیرد.

پاسخ ها ممکن است تحت تاثیر مصاحبه گر قرار گیرد. گروه هایی با علاقی خاص ممکن است داده هایی تولید کنند که تمایل به موقعیت سازمانی شان داشته باشد، در حالی که ممکن است داده های متناقض با آن موقعیت، ضرورتاً در دسترس یا آماده نباشد.

داده ها چگونه جمع آوری شده اند؟ سه شکل اصلی برای جمع آوری داده ها وجود دارد:

خودتکمیلی: در این حالت، افراد فرم های خود را پر کرده و در موقع مقرر می توانند آن را تکمیل نمایند. در این شیوه گردآوری ممکن است محدودیت هایی در زمینه تعداد و پیچیدگی سوالات مورد پرسش ایجاد شود و نیز ممکن است پاسخ ها فاقد جزئیات یا صحت لازم باشند. سرشماری نمونه ای از شیوه ی خودتکمیلی است.

آمارگیری مصاحبه محور: در این شیوه یک پرسش گر با گروهی از پاسخگویان منتخب آمارگیری به صورت حضوری یا تلفنی تماس برقرار می کند. معمولاً این شیوه گردآوری منجر به نرخ پاسخگویی بالاتری می شود ولی احتمال خطر تعصب و سوگیری مصاحبه گر نیز در آن وجود دارد. معمولاً می توان سوالات بیشتر و پیچیده تری پرسید و جواب های جزئی تر و صحیح تری دریافت کرد.

داده های اداری: این داده ها حاصل ترانکس های اداری موسسات و کسب و کارهای دولتی نظیر آمار موالید و مرگ و میر می باشند. استفاده از این نوع داده ها، نیاز به سرشماری و آمارگیری را به حداقل رسانده و موجب صرفه جویی در هزینه ها می شود. با وجود این باید به یاد داشت که اطلاعات برای یک هدف متفاوت گردآوری شده و ممکن

است در هنگام مقایسه ی داده ها از منابع مختلف، مشکلات مقایسه ای به وجود آید.

آیا داده ها برای هدف مورد نظر مناسب هستند؟ کاربران برای استفاده ی آگاهانه از داده ها لازم است بدانند که داده ها چه چیزی را نشان می دهند، چگونه باید تفسیر شوند و به هنگام تفسیر آن ها چه مشکلاتی ممکن است به وجود آید؟

برای تعیین مناسب بودن داده ها، موارد زیر را باید در نظر گرفت:

هدف مورد نظر برای گردآوری داده ها چه بوده است؟

• آیا اطلاعات معرف جمعیت کل هستند؟

• خطاهای استاندارد نسبی به چه میزان هستند؟ در صورت زیاده بودن خطاهای

استاندارد، آیا داده ها قابل اعتماد می باشند؟

• داده ها چقدر جدید هستند؟ آیا آخرین اطلاعات موجود می باشند؟

• آیا به دنبال دریافت یک تصویر لحظه ای هستید یا یک روند در طول زمان؟

• آیا منابع داده ها دیگری به منظور مقایسه موجود هستند؟ آیا مجموعه

داده ها قابل مقایسه می باشند؟

• چه فراداده هایی (بیانیه ی کیفیت، و یادداشت های توضیحی) در کنار

داده ها وجود دارند؟

توانایی درک مفاهیم آماری

قالب های اصلی ارائه ی آماری: جدول، نمودار، نقشه

انواع مختلف میزان ها: درصد، نسبت، نرخ

مفاهیم پیچیده تر آماری: تفاوت میان میانه، میانگین، σ / تفاوت میان

داده های اصلی، داده های روندی و داده های فصلی/ تفاوت میان آمارگیری و

سرشماری

توانایی تحلیل، تفسیر و ارزیابی اطلاعات آماری

سازمان دهی داده ها، ایجاد و ارائه ی نمودارها و جداول و کار با ارائه های

آماري متفاوت: شخص باسواد از لحاظ آماری، باید بفهمد که چگونه داده های

سازمان دهی شده می تواند به تفسیر آن ها کمک کند. جداول و نمودارها

عموماً برای ارائه ی نتایج استفاده می شوند. جداول جزئیات بیش تری را

فراهم کرده و مقادیر واقعی را نشان می دهند، در حالی که نمودارها برای

نشان دادن ارتباطات مفیدترند و روی قالب، شکل و حرکت داده ها

تمرکز می کنند. نمودارها به ویژه در ارائه ی تغییرات در طول زمان مفید

هستند.

توصیف و خلاصه سازی داده های پایه ای: با توجه به وجود حجم

گسترده ای از داده ها، گاهی اوقات پیدا کردن داده های مورد نیاز مشکل است.

تحلیل دقیق، یک گام حیاتی در بیان داستان مهم موجود در داده هاست.

تحلیل بی کیفیت می تواند منجر به تدوین نتایج نادرست و نامناسب شود. بنا

بر این باید مراقب باشید و نسبت به موارد زیر اطمینان حاصل کنید:

• موضوع و داده های مرتبط با آن را درک کرده اید.

• نسبت به داده های مورد استفاده خود نکته سنخ و نقاد هستید.

• قبل از کسب رضایت از رسم تصویری صحیح و واقعی، به دقت آن را بررسی

کرده اید.

دانش قبلی کمک می کند تا درباره ی نوع داده ها چه انتظاراتی ایجاد

شود اما مراقب محدودیت هایی که آن انتظارات می توانند روی نتیجه ی

تحلیل هایتان ایجاد کنند، باشید. نتایج دور از انتظار، ممکن است

هشدار دهنده باشد، اما به همان نسبت از جهت این که چه چیزی را به شما

نشان می دهد نیز حائز اهمیت است. در مورد نتایج تردید و پرسش هایی

را ارائه کنید و به بررسی و تحقیق بیش تری بپردازید تا راضی شوید که تفسیر

صحیحی از اعداد به دست آورده اید. به یاد داشته باشید ممکن است داده ها

لزوماً آن داستان مورد انتظار را به شما نگویند.

استخراج، درک و توضیح داده های ارائه شده به روش های مختلف:

به هنگام مقایسه محتاط باشید. مقایسه نمی تواند بین «سیب» و «پرتقال»

صورت بگیرد، فقط بین «پرتقال» و «پرتقال».

در موارد زیر دقت کنید:

مقایسه ی داده ها از منابع متفاوت

شما باید بررسی نمایید که آیا مجموعه داده‌ها واقعاً قابل مقایسه هستند.

مثال: ممکن است در استرالیا نتایج حاصل از یک سرشماری در خصوص مراقبت مجانی از کودک مستقیماً با آمارگیری تحت عنوان آمارگیری مراقبت از کودک قابل مقایسه نباشد زیرا سن کودکان ممکن است در هر دو یکسان نباشد. دو سازمان در استرالیا ممکن است اطلاعات درباره‌ی بیکاری جمع‌آوری کنند اما داده‌های آن‌ها قابل مقایسه نباشد چون تعریفی که از بیکاری ارائه می‌کنند ممکن است در دو طرح آمارگیری متفاوت باشد. وقوع تغییرات

مجموعه داده‌ها می‌توانند در طول زمان دستخوش تغییر شوند مثل تغییر در طبقه بندی‌ها، جغرافی، اندازه‌ی نمونه، روش شناسی و ...

مثال: کدهای جدید طبقه بندی صنعت در استرالیا و نیوزیلند در سال ۲۰۰۶ جایگزین نسخه‌ی ۱۹۹۳ شدند تا قابلیت مقایسه در سطح بین‌المللی را داشته باشند، بنا بر این مقایسه‌های مستقیم بین این دو نسخه امکان‌پذیر نیست.

تفاوت تعاریف

تعاریف ممکن است بر حسب آمارگیری یا شرایط متفاوت باشند. همواره اطمینان حاصل کنید که از تعریف صحیح استفاده می‌کنید و از محدودیت‌های داده‌ها مطمئن شوید.

مثال: واژه‌ی «کودک» ممکن است بسته به شرایط، معانی مختلفی داشته باشد: زیر ۱۳ سال، زیر ۱۵ سال، زیر ۱۸ سال. باید مطمئن شوید گروه‌های سنی یکسان را مقایسه می‌کنید.

همبستگی اطلاعات

همبستگی به معنای علیت نیست. ارتباط بین داده و یک رویداد ممکن است صرفاً همزمان با «یا» ممکن است دلایل متعدد دیگری وجود داشته باشد» داده‌هایی باشد که فقط یک جنبه از ارتباط را منعکس نمایند.

مثال: تعداد رو به افزایش حملات کوسه‌ها در ساحل دریای شرق استرالیا در ژانویه ۲۰۰۹ ممکن است با رونق خرده‌فروشی محصولات ضد آفتاب مرتبط باشد. این رونق خرده‌فروشی صرفاً با دوره‌ی اوج حمله‌ی کوسه‌ها هم‌زمان بوده اما بعید به نظر می‌رسد که تعداد حمله‌ها با استفاده‌ی رو به رشد از ضد آفتاب‌ها ارتباطی داشته باشد.

نتایج بدون تغییر

تغییر داده‌ها مهم و اغلب اجتناب‌ناپذیر است، بنا بر این فقدان تغییر در نتایج به مرور زمان مشکوک است.

مثال: اگر نرخ بیکاری در طول چند ماه بدون تغییر باقی‌ماند باید مورد بررسی مجدد قرار گیرد که چرا این‌گونه است.

درک موقعیت

از آن‌جا که بسیاری از داده‌ها وابسته به موقعیت و شرایط هستند، داشتن درک خوبی از آن شرایط بسیار مهم است.

مثال: اطلاعات جالبی که از جایی گلچین شده، ممکن است مفهومی کاملاً متفاوت از زمانی داشته باشد که در جای دیگری قرار بگیرد. در دوره‌ی تغییرات تدریجی، برآوردهای نیروی کار نشان می‌دهد که ایالت تاسمانی - ایالتی در جنوب شرق استرالیا - دارای پایین‌ترین نرخ مشارکت در میان تمام

ایالات استرالیا در سال مالی ۲۰۰۷-۲۰۰۸ داشته است. اما در اکتبر ۲۰۰۸ نرخ مشارکت تاسمانی، در بالاترین حد (۶۰٫۹٪) بود. هر دو ادعا کاملاً صحیح هستند اما گزارش دهی‌گزینه‌ای از این داده‌ها می‌تواند همراه‌کننده باشد. حتی آمارهای قابل اعتماد می‌توانند در صورتی که فقط بخشی از داستان گفته شود، تحریف شوند.

توانایی برقراری ارتباط بین اطلاعات آماری و استنباط‌ها داده‌ها چگونه گزارش می‌شوند؟

تبدیل داده به اطلاعات یک مهارت اساسی است و برقراری ارتباط صحیح با اطلاعات آماری برای تصمیم‌گیری اثر بخش، حیاتی است. برای اطمینان از صحت داده‌ها، سواد آماری از ما می‌خواهد نسبت به شیوه‌ی گزارش دهی داده‌ها و قابلیت اعتماد به نتایج حاصل از آن‌ها تردید نماییم. نتایج بد می‌تواند از داده‌های خوب نیز استخراج شود. باید نسبت به موارد زیر دقت نماییم.

استفاده از اعداد خلاصه‌ی پایه‌ی

گاهی اوقات استفاده از اعداد خلاصه‌ی پایه‌ی ای مثل میانگین‌ها می‌تواند گمراه‌کننده باشد.

مثال: اگر چندین خانه در منطقه‌ای به ترتیب ۲۷۵، ۲۹۵، ۳۰۰، ۳۲۵ و ۸۵۰ هزار دلار قیمت‌گذاری شده باشند، استفاده از میانگین برای محاسبه‌ی متوسط قیمت خانه منتج به عدد ۴۰۹ هزار دلار می‌شود. این موضوع یک اثر توری بسیار بالا از ارزش خانه‌های آن منطقه ایجاد می‌کند. در واقع ارزش میانه‌ی ۳۰۰ هزار دلار تصویر درست‌تری از متوسط قیمت خانه‌ها ارائه می‌دهد.

استفاده از نسبت‌ها

استفاده از نسبت‌ها، به ویژه در ارتباط با اعداد کوچک، می‌تواند منجر به نتایج گمراه‌کننده نیز شود.

مثال: طبق آمار موثق، میزان وقوع قتل در یکی از شهرهای دنیا، از سال ۲۰۰۶ تا سال ۲۰۰۷ در حدود ۵۰ درصد افزایش داشته است. با این‌که چنین آماری صحیح است، اما تعداد واقعی قتل‌ها از ۴ به ۶ افزایش پیدا کرد و آن قدر که به نظر می‌رسید افزایش چشمگیری وجود نداشته است.

متغیرهای فصلی

متغیرهای فصلی می‌توانند روی نتایج تأثیرگذار باشند.

مثال: قیمت خرده‌فروشی یک ماه خاص در یک سال ممکن است از سال بعدی کم‌تر باشد. در ظاهر ممکن است این نتیجه‌گیری منطقی باشد که عایدی کسب و کار متحمل ضرر شده است، اما ممکن است این مسئله صرفاً ناشی از تأثیرات یکی از اعیاد بوده که از آن ماه خاص به ماه دیگری در اوایل سال آینده منتقل شده است. برخی از سازمان‌های آماری، به منظور حذف اثرات متغیرهای فصلی، از تعدیل فصلی برای استاندارد سازی داده‌ها استفاده می‌کنند.

نتیجه‌گیری

سواد آماری در اصل توانایی یافتن، دسترسی، استفاده، درک و برقراری ارتباط با داستان نهفته در داده‌ها است. بنا بر این درک صحیح، تفسیر و ارزیابی منتقدانه اطلاعات آماری می‌تواند به تصمیم‌گیری کمک نماید. اهمیت سواد آماری در جامعه‌ی سرشار از اطلاعات ما، به این معنا است که اکنون به توانایی اصلی نظیر خواندن و نوشتن تبدیل شده است.

آمار روی تمام جنبه‌های زندگی ما نفوذ کرده و تأثیرگذار است و از طریق رسانه‌ها و تبلیغات ما را متقاعد می‌کند که با نقطه‌نظرات خاصی موافقت نماییم و دست به اقدام خاصی بزنیم. پس برای اینکه آمار را خوب بفهمیم و بتوانیم آمارها را به طور مؤثر و صحیح استفاده و تفسیر کنیم، باید سواد آماری داشته باشیم.

منابع و مأخذ:

لزوم سواد آماري در مدیریت

سیده محدثه حجتی و الهه آزرمی

تهیه کنندگان :



نکات فنی و فرمولهای ریاضی آن است؟ اینکه مدیری بداند توزیع نرمال دارای چه فرمولی است در کجای کار عملی و مدیریت از آن استفاده خواهد کرد؟ مبحث سواد آماری، بیشتر از آنکه به چنین مباحثی بپردازد به درک درست مدیران از آمارهای مدیریتی اهمیت میدهد، چرا که آنها یکی از مصرف کنندگان اصلی آمار و اطلاعات هستند.

نکته مهمی که برنامه ریزان در تفکر آماری می توانند مورد توجه قرار دهند تفاوت نگاه آمار به اطلاعات نسبت به تفکر غیر آماری است. علی رغم ارتباط قوی این علم با علوم دیگر مانند ریاضی، باید گفت که اطلاعات آماری ارتباط صددرصد با مفهوم و متن ارائه شده در گزارش آماری دارد و اعداد و ارقام آماری، به تنهایی نشان دهنده چیزی نخواهد بود. به طور دقیق تر، مباحث آماری به صورت مجرد، قابل تجزیه و تحلیل نیستند. برای مثال گرچه ده درصد یک عدد قابل قبول برای انجام محاسبات ریاضی است ولی به تنهایی در تفکر آماری مفهومی ندارد و این سوال را پیش می آورد که ده درصد چه چیزی یا چه موضوعی؟ در حقیقت بیان اعداد در یک مفهوم ارائه شده نیاز به توانایی استنباط و برقراری ارتباط بین آن اعداد و مفهوم مورد اشاره دارد.

شاید یک مدیر یا فرد آماردان در خدمت مدیریت یک سازمان، بتواند تمامی محاسبات لازم برای به دست آوردن اطلاعات آماری را انجام دهد، ولی بیان درست و استنباط صحیح از آمار ارائه شده چیز دیگری است. برای مثال درک این مطلب که ارتباط لزوماً رابطه علت و معلولی را نشان نمی دهد ولی برای مطالعه یک رابطه علت و معلولی دانستن مقدار ارتباط می تواند راهگشا باشد، نکته ای است که در فرمولهای همبستگی توانایی نشان دادن آن وجود ندارد. شنیدن این مطلب که کشوری بهترین رشد ناخالص ملی و سطح بهداشت را در منطقه خود دارد، شاید برای فردی که تفکر آماری ندارد جالب باشد، ولی یک مدیر با سواد آماری که لزوماً متخصص رشته آمار نیست در لغت «بهترین» کنجکاو کرده و اختلاف معنی دار با کشورهای بعدی را جستجو خواهد کرد چرا که ممکن است بهترین کشور لزوماً اختلاف معنی داری با رقبای خود نداشته باشد. این مطلب به تسلط کامل بر فرمول های اختلاف معنی داری نیاز ندارد بلکه آنچه نیاز دارد تفکر آماری است.

همانطور که می دانیم فرضیات آماری می تواند نتایج یک بررسی آماری را تحت تاثیر قرار دهد. برای مثال اگر نمونه مورد بررسی را حداقل نمونه لازم در نظر بگیریم می توانیم امیدوار باشیم اختلافات بین گروه ها بیشتر نشان داده شود و یا برای کاهش اختلافات بین گروه های مورد بررسی شاید با افزایش حجم نمونه به مقصود بهتر برسیم. یک مدیر دارای سواد آماری می داند که در نظر گرفتن فرضیات ابتدایی یک بررسی آماری بر اساس داده چقدر میتواند بر روی درصدهای به دست آمده از یک مطالعه آماری تاثیر بگذارد. مخصوصاً زمانی که فاکتورهای مرتبط و مداخله گر وجود دارند. برای مثال یک مدیر ارشد ممکن است رئیس یک بخش را بر اساس پایین بودن نرخ کارایی بازخواست کند حال آنکه به وضعیت نامناسب محیطی و عوامل تاثیرگذار بیرونی دقت نکرده است. یک مدیر دارای سواد آماری توانایی شناخت متغیرهای مداخله گر و فرضیات آماری را کسب کرده و در تفسیر اخبار و گزارش های آماری به آن توجه دارد.

نکته بعدی توجه به تفکر ارزیابانه است. مدیران باید دارای قابلیت درک مقدار حمایت شواهد آماری از یک نتیجه گیری یا تصمیم مدیریتی باشند. برای یک مدیر، شناخت اعداد و یا انجام سریع محاسبات ذهنی لزوماً مهم تلقی نمی شود که شاید این توانایی در یک فرد عادی نیز باشد. بلکه نیاز به هنر ارزیابی اعداد به عنوان شواهد آماری مطرح است و اینکه در استنباط نتایج یک تحقیق آماری بتواند مطالب ارائه شده را به طور صحیح و واضح تحلیل کرده و به عنوان مثال متوجه شود که ارزیابی ایجاد شده در نتایج یک تحقیق دارای اهمیت بیشتری از تصادفی بودن یا نبودن نمونه ها است و نتایج را بیشتر تحت تاثیر قرار می دهد

حتی قوی ترین مدیران نیز نمی توانند ادعا کنند که بدون استفاده از علم آمار، اطلاعات کافی از تمامی جوانب سازمان داشته و می توانند بهترین راه حل ها را برای حل مشکلات ارائه نمایند. اگر بپذیریم در هر سطحی از سازمان، ممکن است داده هایی مفید ولی دور از چشم همگان مانند تصویر گودال آب وجود داشته باشد، طبیعی است که باید به دنبال راه هایی برای شناسایی این داده ها و تبدیل آنها به اطلاعات مفید باشیم، تا با بکارگیری این اطلاعات در رویه تصمیم گیری بتوانیم تصمیماتی دانش محور گرفته و مدیریتی خردمندانه را ارائه نماییم. مطالعات آماری در سازمان و مخصوصاً حوزه منابع انسانی، در جهت شناسایی محیط خارجی و داخلی سازمان و بهبود مستمر در تمام ابعاد به مدیران کمک میکند.

تمامی سازمانها برای حفظ و ارتقاء پیوسته و بهبود مستمر، نیازمند سه عنصر اساسی دانش، خلاقیت و بهره وری می باشند. یکی از سیستم هایی که باعث ارتقای دائمی و مستمر این سه عنصر میشود، علم آمار و استفاده از روشهای آن است. ابزاری به منظور زدودن غبار نادانی از فضای تصمیم گیری و دوری از رفتارهای بدون دانش است.

علم آمار حاوی روش هایی برای فعال نمودن توان فکری مدیریت در ایجاد بهبودهای مستمر در کارهای مربوط به خود است. بهبودی مستمر که میتواند پروانه ای زیبا و سبک بال را از کرمی زشت و سنگین بیافریند. اجازه دهید بحث را بازتر کنیم:

برخی از مدیران علی رغم اینکه یکی از مصرف کنندگان اصلی آمار و اطلاعات هستند ترجیح میدهند به جای توجه به نیاز روزانه خود به آمار به موقع و تفسیر درست آن، تنها به دانش مدیریت تکیه کنند. پر واضح است که مدیران به استفاده و تفسیر داده های آماری، برای برنامه ریزی و گرفتن تصمیمات درست و به موقع نیازمندند. شاید به علت همین است که اول آبان ماه هر سال روز آمار و برنامه ریزی نامگذاری شده است و این نامگذاری نشانگر اهمیت موضوع است. اضافه کردن مباحث سواد آماری به مباحث لازم علم مدیریت، کمک خواهد کرد تا مدیران تفکر سیستمی و درستی در ارتباط با داده های حوزه مدیریت خود پیدا کنند. سواد آماری برای مدیران به دور از مباحث تئوریک سنگین آمار و احتمالات، بر روی درک صحیح و استفاده به موقع از داده ها و آمارهای ارائه شده تمرکز میکند تا قابلیت درک روابط و شواهد واقعی محیط کسب و کار را به مدیران اهدا کند. دوری از مباحث صرفاً تئوریک آمار ممکن است در نگاه اول، مباحث سواد آماری برای مدیران را ساده نشان دهد ولی در واقع ایجاد یک درک صحیح، عمیق و کاربردی در زمینه آمار به هیچ وجه کار ساده ای نیست و نیاز به تلاش فراوان توسط متخصصان این حوزه دارد. به قول معروف سخن ساده گفتن، سخت است.

به نظر می آید اطلاع از مباحث ریاضی و محض علم آمار و احتمال نباید یک الزام برای مدیران تلقی شود. در حقیقت وقتی می گوئیم مدیران در سراسر دنیا به آمار نیاز دارند سوء تفاهمی صورت میگیرد و از آمار به مباحث تئوری علم آمار و احتمال برداشت میشود. بنابراین مباحثی مانند احتمالات، نمونه گیری و خطاهای آن، توزیع های آماری، فاصله اطمینان، آزمون فرض، معنی داری، مدل سازی، پیش بینی و مباحثی از این قبیل برای آنها تجویز میشود. مباحثی که شاید کاربرد شایسته ای از آمار را برای مدیران متصور نمیکند، گرچه مبانی آمار و تفکر احتمالی و غیر حتمی را پیش روی آنها قرار میدهد و برداشتی از داده را با درصدی خطا میسر میسازد.

برای افرادی که هر دو محیط علمی و عملی آمار را تجربه کرده اند پرواضح است که علم آمار ارائه شده به مدیران با آنچه در عمل با آن سر و کار دارند نیاز به همخوانی قوی و خوبی دارد. به راستی برای ایجاد درک صحیح خطاهای آماری، معنی داری، فواید تصادفی بودن نمونه و غیره برای مدیران چه نیازی به بیان کلیه

یا خیر.

موضوع دیگر تفکر قضاوت گرانه است. این قدرت تفکر در یک مدیر به او این امکان را می دهد که در مورد اینکه چه چیزی را باور کند یا بر اساس داده های آماری انجام دهد، تصمیم درست بگیرد. در حقیقت او می داند چرا این تصمیم بهترین است و دلایل آن چیست. این امر محقق نمی شود مگر اینکه وی با مفاهیمی مانند هزینه_فایده و ارتباط آن با احتمال و ریسک در هنگام تصمیم گیری آشنا باشد و نتایج آماری ارائه شده در این زمینه ها را تحلیل کند.

مطالب ارائه شده، مفاهیمی از علم آمار است که کارکنان سازمان ها، مخصوصا مدیران، با این مطالب درگیر هستند تا بتوانند تصمیمات درست را بر اساس درک صحیح از آمارهای ارائه شده، برای سازمان خود به ارمغان آورند. شاید شایسته باشد این را سواد آماری برای مدیران بنامیم و با توجه به اینکه سال ۲۰۱۳ را سال جهانی آمار می دانیم، جهت گسترش بیش از پیش آن از خود سوال کنیم « چه اطلاعات آماری در محیط داخلی و خارجی سازمان، رسانه ها و جامعه برای کارکنان و مخصوصا مدیران و برنامه ریزان وجود دارد که باعث میشود برنامه ریزی و آمار، در یک روز ملی، کنار یکدیگر قرار داده شوند؟» سواد آماری برای مدیران، نیازهای آماری آن ها را بر اساس آمارهایی تعریف می کند که هر روز در فعالیت مدیریتی خود با آن سر و کار دارند و مطمئنا هدف آن ایجاد یک آماردان علمی و تئوریک نیست.

حجم داده های آماری روز به روز در حال افزایش است و در دنیای کنونی، جنس اطلاعات آماری از داده های به دست آمده بر اساس آزمایش ها، به داده های حاصل شده از مشاهدات تغییر پیدا کرده است. مباحث نسبتا جدیدی مانند داده کاوی و تحلیل منابع داده ای حجیم در کنار مباحث سنتی مانند طرح آزمایش ها و نمونه گیری مطرح شده اند و مطالعات قبل و بعدی، جای خود را به مطالعات چندسطحی و داده های پنل، فضایی، سانسور شده، رکورد و غیره واگذار کرده است. این تغییرات را هنگامی که در کنار نیاز به تصمیم گیری بر اساس اطلاعات و تحلیل های مستند و علمی بگذاریم، عمق قضیه و نیاز روز افزون به افزایش دانش در زمینه آمار مشخص می شود.

سوال اصلی این است. آیا لازم است تا یک مدیر خردمند، توسط تحلیل درست موقعیت و انجام واکنش به موقع بر اساس بررسی های علمی و صحیح، پاسخگوی نیازهای واقعی حوزه مدیریت خود باشد؟ اگر پاسخ مثبت است، منطقی است که مدیران، راهبرد تقویت سواد و تفکر آماری را بیش از پیش در حوزه مدیریتی خود تقویت کنند و نام آن چیزی را که بر اساس تجربه انجام می دهند را بطور علمی نیز بشناسند.

در دنیای امروز چنین به نظر می رسد که گزارش های مدیریتی نیازمند مباحثی عددی و استنباطی از علم آمار هستند که مدیران را به داشتن تفکر آماری بیش از پیش سوق می دهد. مباحثی که با تجربیات و نیازهای روزانه آنها منطبق بوده تا بتوانند مسائل خود را به طور کاملا کاربردی در مباحث طرح شده

بینند. در حقیقت در دنیای اطلاعات و دانش امروزی، همراهی مثلث گونه مدیر، تصمیم گیری و سواد آماری، کلید ایجاد مدیریت کاربردی و موثر است. به نظر نویسنده رعایت اصول سواد آماری در یک گزارش مدیریتی می تواند درک مطالبی از قبیل این مطلب که برخی از نتایج فرایندهای تصادفی موجود در دنیا در تعداد مناسب وقایع مفهوم پیدا می کنند را سهل کند. در حقیقت زمانی که تعداد محدودی منابع اطلاع در رابطه با یک فرایند تصمیم گیری در اختیار است، سخت می توان به اطلاعات صحیح مورد انتظار دست پیدا کرد و باید به تعداد مطلوبی از داده های صحیح، به هنگام و مرتبط دسترسی داشت که به صورت علمی گرد آوری شده اند. شناخت جایگاه مطالعات آماری در بررسی های محیط خارجی و داخلی سازمان جهت تشخیص وضعیت موجود و مطلوب در راستای تمام برنامه ریزی های کلان و راهبردی تا جزئی و همچنین تهیه شاخص های مناسب جهت کنترل و ارزیابی وضعیت و میزان احراز شدن اهداف برنامه برای یک مدیریت موثر می تواند حیاتی باشد. توانایی تشخیص یک مطالعه آماری خوب از یک مطالعه ناکامل و آشنایی کلی با کارکردها و نقاط قوت و ضعف روش های مختلف تحلیل داده ها، بدون نیاز به فرمول های ریاضی یا مبانی محض احتمالی آن ها، پایه و اساس کار است که تمرکز بیشتر بر روش های آماری مانند شاخص سازی، مدل سازی، مدیریت ریسک و پیش بینی آینده بر اساس علم آمار و مدل های ارائه شده را می طلبد. برای مثال آشنایی با کارکرد روش شبیه سازی مونت کارلو یا انواع روش های دیگر شبیه سازی در تحلیل ریسک و نحوه تفسیر نتایج مربوط به آن در عصر حاضر کاری پیچیده و دور از انتظار نیست. درک رابطه شیوع و بروز در روابط بیان شده نوسط آمار، همانقدر برای یک مجموعه مدیریتی لازم است که درک رابطه سهام و بازده در اقتصاد یا بدهی و کسر بودجه در حسابداری لازم است. همانطور که می دانیم ممکن است در یک زمان مشخص و ثابت نرخ بیکاری در یک تجارت خاص قابل کاربرد و تکرار باشد ولی شیوع بیکاری در طول زمان، نرخ افزایش بیکاری را تغییر داده و نتایج با گذشت زمان تغییر خواهد کرد و در حقیقت شیوع یک واقعه، در نتیجه رفتار بروز یافته کنونی، ممکن است نتایج آینده را متفاوت سازد. درک نیاز و فایده بررسی آماری برخی گروه ها و وقایع به صورت جداگانه، مانند درک الزام بررسی جداگانه ریسک سازمان های قدیمی و جدید و یا بررسی آماری شکایات نه تنها بر اساس تعداد آنها بلکه بر اساس شدت تاثیر آنها در اعتبار سازمانی مانند برخی مفاهیم حسابداری است که نیاز به بررسی جداگانه هزینه های ثابت و متغیر، قابل کنترل و غیرقابل کنترل و مستقیم و غیرمستقیم دارند و این الزام باید در یک گزارش مدیریتی مطلوب در نظر گرفته شود.

منابع مأخذ:

۱. <http://www.SavadeAmari.com>

۲. <http://www.irstat.ir>

آغاز يك مسير: اتحاديه دانشجويي آمار ايران

بنيامين لقمان نيا

تهيه كننده:



انجمن علمی دانشجویی امکانی برای تمرین داوطلبانه فعالیت‌های علمی در قالب کار گروهی است. دانشجویان با اهداف علمی و تخصصی گرد هم می‌آیند و با فعالیت در انجمن‌های علمی ضمن ترویج و تعمیق فضای علمی در دانشگاه‌ها، زمینه رشد خود را نیز فراهم می‌آورند. انجمن علمی ضمن آن که امکانی است برای تعمیق و ترویج علم در فضای دانشگاه، فرصتی برای پیاده‌سازی آموخته‌ها و مشاهده نتایج آن در عمل نیز هست.

بنابراین دور از انتظار نخواهد بود که دانشجوی فعال در انجمن‌های علمی را فردی بدانیم که به مراتب نسبت به هم‌کلاسی‌های خود توانمندتر و باتجربه‌تر است؛ توانمندی‌هایی از جمله:

- مهارت کار گروهی
- مهارت تبدیل ایده به عمل
- مهارت ارتباط
- مهارت مدیریت و اجرا
- مهارت تامین هزینه‌های مالی
- مهارت هم‌اندیشی، آزاداندیشی و تفکر انتقادی.
- انجمن‌های علمی دانشجویی در چارچوب آیین‌نامه خود، قادر به برگزاری همایش، سمینار، نشست علمی، سخنرانی، کلاس و کارگاه‌های آموزشی مورد نیاز، بازدیدهای علمی مرتبط با هر رشته، کرسی‌های آزاد اندیشی با موضوعات مرتبط با هر رشته، همایش‌ها و نشست‌های ملی، برگزاری مناظره‌های دانشجویی، برگزاری مسابقات علمی، شرکت در مسابقات علمی، انجام فعالیت‌های آموزشی و پژوهشی و... هستند.

نخستین نمونه از انجمن‌های علمی دانشجویی از سال ۱۳۶۳ شروع به کار نموده‌اند. با تصویب اولین آیین‌نامه رسمی تأسیس و فعالیت انجمن‌های علمی دانشجویی در سال ۱۳۷۸، ساختار منسجمی در درون معاونت فرهنگی و اجتماعی وزارت علوم، تحقیقات و فناوری برای این تشکلهای به وجود آمد که امروزه پس از گذشت حدود دو دهه از آغاز فعالیت رسمی انجمن‌ها، بیش از ۹۵۰۰ انجمن علمی دانشجویی در دانشگاه‌های تابع وزارت علوم شکل گرفته است (واحدهای دانشگاه آزاد اسلامی از این آمار مستثنی هستند) که در حال سازماندهی شبکه‌ای بسیار عظیم از دانشجویان در رشته‌های تخصصی می‌باشد. به دنبال افزایش یافتن انجمن‌های علمی در هر رشته، ایده تأسیس اتحادیه‌های تخصصی نیز شکل گرفت که تاکنون ۳۶ اتحادیه از درون این انجمن‌ها تأسیس شده و این روند همچنان ادامه خواهد داشت. با تمهیدات در نظر گرفته شده از سوی وزارت علوم، فعالیت انجمن‌ها و اتحادیه‌های دانشجویی به مثابه ظرفیتی برای حضور نخبگان علمی در راستای تحقق راهبرد «سازماندهی و تقویت انجمن‌ها و جمعیت‌های علمی به منظور ایفای نقش مرجعیت علمی و ارتقای مشارکت در تصمیم‌سازی‌ها و توسعه، ترویج و انتشار علم و فناوری» در سند «نقشه جامع علمی کشور» و نیز اقدامات ۲ و ۳ راهبرد ملی ۲ از راهبرد کلان ۲ در «سند راهبردی کشور در امور نخبگان» مصوب سال ۱۳۹۱ شورای عالی انقلاب فرهنگی قرار گرفته است.

اتحادیه انجمن‌های علمی دانشجویی آمار در آبان سال ۱۳۹۵ با دغدغه رشد و اعتلا، و در راستای ایجاد انسجام در روند فعالیت انجمن‌های این رشته تأسیس شد. اتحادیه، سازمانی غیرانتفاعی متشکل از انجمن‌های علمی دانشجویی فعال در حوزه‌های علمی مرتبط با رشته‌ی آمار که در دانشگاه‌ها و مؤسسات آموزش عالی کشور فعالیت می‌کنند، است. از اهداف تأسیس و فعالیت اتحادیه می‌توان ایجاد زمینه مناسب برای تجلی

استعدادهای و برانگیختن روحیه مشارکت جمعی و خلاقیت و شکوفایی علمی دانشجویان رشته‌ی آمار، ارتقای سطح فعالیت انجمن‌های علمی از طریق هم‌افزایی و همکاری مشترک و تکثیر ایده‌های موفق، ایجاد زمینه همکاری با مراکز علمی، پژوهشی یا صنعتی، ایجاد شبکه ارتباطی میان دانشجویان آمار سراسر کشور و تعامل مستمر با آنان و نیز توانمندسازی علمی و کاربردی دانشجویان رشته‌های آمار جهت آماده‌سازی برای ورود به بازار کار اشاره کرد.

اتحادیه دانشجویی آمار ایران نمایندگی علمی دانشجویان تمام گرایش‌های رشته آمار را بر عهده داشته و در حوزه‌هایی نظیر ارتباط با نهادهای موثر در حوزه سیاست‌گذاری آموزشی و پژوهشی برگزاری، همکاری در برگزاری گردهمایی، کنفرانس و سمینارهای علمی در سطح ملی، برگزاری مسابقات علمی در سطح ملی، برگزاری کارگاه‌های آموزشی تخصصی، انتشار نشریات سراسری و ... دنبال خواهد کرد. حوزه‌های آموزش، پژوهش و ارتباط با صنعت به صورت همزمان در دستورکار اتحادیه قرار دارد و برنامه‌های مصوب اتحادیه شامل هر سه حوزه می‌باشد. در طراحی این برنامه‌ها، مهارت‌افزایی، کشف استعداد و سپس کمک به تقویت مهارت‌ها و سرانجام برقراری ارتباط با صنعت، مورد توجه شورای مرکزی اتحادیه می‌باشد. مهمترین برنامه‌های پیش‌روی اتحادیه به شرح ذیل است:

- نخستین کنفرانس ملی دانشجویی آمار ایران: با توجه به اینکه هر ساله کنفرانس علمی آمار تحت حمایت انجمن آمار ایران برگزار می‌شود و اکثریت شرکت‌کنندگان دانشجویان مقاطع کارشناسی ارشد و بالاتر هستند، انجمن علمی دانشجویی آمار دانشگاه تربیت مدرس زیر نظر اتحادیه انجمن‌های آمار در نظر دارد با هدف مشارکت دانشجویان مقطع کارشناسی نیز، راه و انگیزه آنان در ارتقا سطح علمی خود را به وسیله اعطای امتیاز و جوایز نفیس افزایش دهد. همچنین این طرح سبب افزایش مهارت دانشجویان مقطع کارشناسی در انتخاب زمینه مورد علاقه آن‌ها با دیدی وسیع و عمیق در مقطع بالاتر خواهد بود.

- نخستین دوره لیگ مسابقات دانشجویی آمار ایران: با وجود این که مسابقات دانشجویی آمار ایران توسط انجمن آمار ایران به صورت سالانه برگزار می‌شود اما اتحادیه آمار از ابتدای فعالیت خود به دنبال آن بوده تا با ایده‌های نوآورانه، طرحی نو در انداخته و ضمن برگزاری یک مسابقه جذاب، حلقه ارتباط با صنعت را نیز شکل دهد. این مسابقه در دو سطح دانشجویان کارشناسی و دانشجویان کارشناسی ارشد و نیز به دو صورت فردی و گروهی برگزار خواهد شد. فرآیند مسابقات لیگ از سه هفته تشکیل شده است که بین هر کدام از هفته‌ها، یک هفته برای استراحت در نظر گرفته شده است. فرآیند اجرای هفته‌های لیگ، به صورت همزمان و جداگانه در هر دانشگاه اما فینال مسابقات به صورت متمرکز در یک محل انجام خواهد شد. سؤالات طرح شده، قابل حل توسط نرم‌افزارهای آماری بوده و شرکت‌کنندگان مجاز خواهند بود که از نرم‌افزارهای مصوب مسابقه که بر روی سیستم‌های مسابقه نصب شده است، به تناسب سلیقه و نیاز خود استفاده کنند.

- نخستین گردهمایی ملی شرکت‌های آماری: اتحادیه در نظر دارد با حمایت انجمن آمار ایران، نخستین گردهمایی شرکت‌های فعال در حوزه آمار را با موضوع «آمار و پژوهش» در آذرماه ۱۳۹۷ به مناسبت «هفته پژوهش» و به میزبانی دانشگاه گیلان برگزار نماید. شرکت‌های آماری، شرکت‌های مشاوره‌های هستند که- صلاحیت آنها از سوی مرکز آمار ایران بررسی و تأیید شده و در زمینه مشاوره، نظارت و اجرای طرح‌های آماری مورد نیاز سازمان‌ها، شرکت‌ها، مراکز آموزشی و پژوهشی، اساتید، دانشجویان و پژوهشگران رشته‌های

مختلف، فعال هستند. از اهداف این گردهمایی دو روزه می‌توان به ایجاد فضای تبادل تجربه مشاوران و شرکت‌های آماری فعال در سطح کشور، ایجاد انگیزه در دانشجویان رشته آمار از طریق آشنایی با فعالیت‌های شرکت‌های آماری و داده‌پردازی و نیز فراهم آوردن فضایی برای آشنایی شرکت‌های آماری با نیازهای پژوهشی سازمان‌ها، شرکت‌ها و پژوهشگران حوزه‌های مختلف اشاره کرد.

هر آغازی، ناگزیر از یک پایان است. از همه بزرگوارانی که در طول دو سال فعالیت در انجمن و سپس دو سال در سطح اتحادیه، این عضو از خانواده بزرگ آماری‌ها را یاری و همراهی نموده‌اند صمیمانه سپاسگزارم. همچنین به عزیزانی که قصد آغاز همکاری با انجمن علمی و اتحادیه داشته باشند، پیشاپیش تبریک و خوش‌آمد گفته، دستشان را به گرمی خواهیم فشرد و در ایام باقیمانده، در کنارشان خواهیم بود. امروز در قدم‌های پایانی مسیری قرار دارم که از چهار سال قبل آغاز شده بود؛ چهار سال کنکاش برای گشودن مسیری به آینده که گاهی چشیدن شیرینی کامیابی‌ها و اوقاتی هم تحمل تلخی ناکامی‌ها بوده‌است. برگ‌های این دفتر نیز رو به اتمام است اما صفحات سختی که در این دفتر با تلاش و امید سپری شده بود، نوید روزهای درخشان را می‌دهد تا حکایت همچنان باقی بماند...

معرفی نرم افزارهای آماري (۲)

صادق بخشی پور

تهیه کننده:



در شماره اول نشریه کارا که در پائیز ۱۳۹۶ توسط انجمن علمی آمار دانشگاه گیلان منتشر شد، در بخشی از این نشریه به معرفی برخی از نرم افزارهای پرکاربرد آماری (R, SPSS, MINITAB, SAS) پرداخته شد. حال تصمیم گرفته ایم، مختصراً برخی دیگر از نرم افزارهای دنیای آمار را به شما معرفی کنیم.

• نرم افزار Stata

Stata (استتا) یک برنامه آماری چند منظوره است که اولین نسخه آن در سال ۱۹۸۵ عرضه شد. این برنامه بین دانشگاهیان و موسسه‌های آکادمیک سراسر دنیا کاربرد فراوان دارد. اکثر کاربران این نرم افزار محققینی می‌باشند که در رشته‌های اقتصاد، جامعه شناسی، علوم سیاسی و شاخه‌هایی از علوم پزشکی مشغول به پژوهش و تحقیق می‌باشند. از جمله قابلیت‌های آن می‌توان به مدیریت داده‌ها، تجزیه تحلیل آماری، ترسیم نمودار، شبیه سازی و برنامه نویسی به صورت دلخواه اشاره نمود. کلمه Stata ترکیبی از دو کلمه Statistics و Data می‌باشد و تلفظ آن معمولاً به شکل خاصی نمی‌باشد. استثناً اغلب روی یک رابط نوشتاری که دستورها را باید در قالب متن برای اجرا در نرم افزار نوشت تأکید می‌کند چرا که اعتقاد دارد این روش باعث افزایش سرعت می‌شود. با این حال از نسخه ۸ نرم افزار به به بعد یک واسط گرافیکی به شکل منوی افقی در بالای نرم افزار در دسترس می‌باشد که بعد از کلیک روی عملیات مورد نظر یک جعبه تبادلی دوطرفه برای اجرای دستور ظاهر شده که شما را برای اجرای دستور راهنمایی می‌کند. خصوصیت خوب نرم افزار این است که پس از اجرای دستور از طریق رابط گرافیکی، دستور نوشتاری آن نیز نمایش داده می‌شود تا کاربر با فراگرفتن آن بتواند در مرحله بعدی کار را با سرعت بیشتری انجام دهد. دستورهایی که از قبل اجرا شدند را نیز با یک کلیک ساده روی پنجره دستورات میتوان دوباره اجرا کرد. همچنین با ساخت یک Dofile می‌توانید تمامی دستورات مورد نظر را به یکباره اجرا نمایید.

* قابلیت های کلیدی نرم افزار Stata:

۱) مدیریت داده‌ها

۲) قابلیت برنامه نویسی

۳) تجزیه و تحلیل آماری

۴) امکان شبیه سازی داده‌ها

۵) نمایش گرافیکی و نموداری داده‌ها

۶) ذخیره‌ی داده‌ها به صورت صفحه گسترده

۷) امکان وارد کردن اطلاعات از نرم افزار اکسل

۸) قابلیت خواندن داده‌های با فرمت اسکی (ASCII)

۹) رابط کاربری ساده جهت افزایش سرعت عملکرد برنامه

• نرم افزار Eviews

نرم افزار Eviews (مخفف عبارت Econometric Views) یکی از نرم افزارهای تخصصی در زمینه بکارگیری روش های اقتصادسنجی به ویژه روشهای جدید آن می‌باشد. چنانچه محقق اقتصادی بخواهد بطور جدی در زمینه تحقیقات اقتصادی فعالیت نماید ضروری است در کنار آموزش تکنیک های جدید اقتصادسنجی با نرم افزارهای با قابلیت نظیر Eviews نیز آشنا شوند. با توجه به گسترده شدن روزه روز علم اقتصاد و مدل‌های اقتصادی و پیچیدگی روابط بین متغیرهای اقتصادی دیگر به هیچ عنوان امکان تحلیل و تخمین سیستم ها و مدل‌های اقتصادی بدون کمک گرفتن از نرم افزارها میسر نیست.

نرم افزار Eviews جهت آنالیز داده‌های متداول آماری و اقتصادی مانند: (آنالیز داده‌های پانل) و (تخمین داده‌های سری زمانی) کاربرد دارد و از فرمول‌های محاسباتی و تکنولوژی ارتباط بین داده‌ای همراه با تمرین‌های متداول ساده به

صورت یک پکیج نرم‌افزاری تهیه گردیده است.

شما می‌توانید به راحتی با استفاده از نوار ابزار و منوی این نرم افزار و با کمک گزینه‌های راهنما از آن برای توابع مورد نظر خود استفاده نمایید و به این ترتیب با باز کردن هر پنجره می‌توانید اطلاعات لازم را برای فرمان مورد نظر خود بدست آورید. از قابلیت های نرم افزار Eviews این است که می‌تواند داده‌ها، ورودی و خروجی‌ها، نمودارها و جداول را سازماندهی کند. در این نرم افزار هر کدام از این داده‌ها می‌توانند، کپی، ذخیره و یا کات شوند و در پنجره‌ای دیگر مورد استفاده قرار بگیرند.

* ویژگی ها و امکانات نرم افزار Eviews:

۱) تخمین های خطی و غیر خطی سیستم معادلات ساده و سیستم معادلات

همزمان

۲) محیط کاربری آسان و رابط گرافیکی و کاربری خوش ساخت

۳) امکان تبادل اطلاعات با دیگر نرم افزارهای تحت ویندوز

۴) محاسبه سری های زمانی و ایجاد آنها

۵) توانایی پذیرش، خوانش و نوشتن حجم گسترده‌ای از فرمت‌های گوناگون داده‌ها، بدون واسطه

۶) ایجاد بانک اطلاعاتی

۷) طراحی شده بر مبنای مفهوم آبجکت، با مؤلفه‌های منحصر بفرم

۸) رسم نمودارهای پراکنش دایره ای و میله ای و نقطه ای و دیگر نمودارهای کاربردی

۹) امکان تخمین و شبیه سازی الگوهای اقتصادسنجی با بهترین نتایج و تخمین

مدل های اقتصادی با روش های VAR و VECM

۱۰) قابل استفاده برای همه، بدون نیاز به آشنایی با زبان‌های برنامه نویسی

۱۱) برخورداری از مجموعه ابزارهای نوین اقتصادسنجی و آماری کامل و کارآمد، مطابق با متدهای جدید و استاندارد

۱۲) امکان برنامه نویسی در محیط نرم افزار برای تحلیل سیستم های کلان اقتصادی

• نرم افزار Statistica

اگر احتیاج به رسم نمودارهای آماری با کیفیت بالا جهت چاپ دارید یا نرم افزارهای دیگر نمی‌توانند نمودار مورد نظر شما را رسم کنند حتماً این نرم افزار را امتحان کنید. نرم افزار Statistica محصولی از شرکت StatSoft می‌باشد و اخیراً نسخه ۱۰ آن منتشر شده است. این نرم افزار تا نسخه ۷ بصورت Crack شده موجود است اگرچه در نسخه کرک شده قابلیت های داده کاوی نرم افزار کاملاً غیر فعال است.

این نرم افزار مورد توجه بسیاری از شرکت های بزرگ صنعتی در دنیاست و قابلیت هایی را داراست که نرم افزار های دیگر از آن بی بهره هستند مثلاً می‌توان از نسخه تحت شبکه این نرم افزار نام برد یا امکان آنالیز همزمان چند کلاینت بر روی یک سری داده و امکان انجام محاسبات تحت وب و با استفاده از اینترنت و یکپارچه سازی دیتابیس های مختلف در نقاط مختلف دنیا ب روی سرورهای متفاوت یک شرکت.

همچنین قسمت Spreadsheet این نرم افزار دارای قابلیت بالا برای کار با حجم زیادی از متغیرها و رکوردهاست. درحالیکه با افزایش تعداد متغیرها در اکثر نرم افزارها سرعت نرم افزار و انجام تحلیل ها بشدت کاهش می‌یابد این نرم افزار براحتی با حجم وسیع داده ها کار می‌کند. برای مقایسه سعی کنید یک متغیر با اعداد تصادفی به تعداد ۱۰۰۰۰ را در این نرم افزار و یک نرم افزار دیگر تولید کنید. این نرم افزار بر اساس Visual Basic نوشته شده و اگر VB بدانید می‌توانید به

گردیده است. اما این روش را می‌توان به عنوان موارد خاصی برای روش‌های تحلیل عاملی تأییدی، تحلیل رگرسیون چند متغیری، تحلیل مسیر، مدل‌های اقتصادی خاص داده‌های وابسته به زمان، مدل‌های برگشت پذیر و برگشت ناپذیر برای داده‌های مقطعی / طولی، مدل‌های ساختاری کوواریانس و تحلیل چند نمونه ای (مانند آزمون فرضیه‌های برابری ماتریس کوواریانس‌های برابری، ماتریس همبستگی‌ها، برابری معادلات و ساختارهای عاملی و غیره) نیز به کار برد.

• نرم افزار AMOS

AMOS مخفف Analysis of moment structures می‌باشد. این نرم افزار برای کاربرد ساده طراحی شده است و خصیصه اصلی آن این است که مدل سازی معادله ساختاری را به شیوه ای ترسیمی ارائه می‌دهد به نحوی که می‌توان به سرعت مدل‌ها را تعریف، محاسبات را انجام و در صورت نیاز آنها را به سادگی اصلاح کرد. در این نرم افزار جعبه ابزاری وجود دارد که به کاربر در ترسیم انواع نمودارهای یاری رسانده و به خوبی امکان تغییر و اصلاح مدل ترسیم شده را فراهم می‌آورد.

AMOS در ابتدا با این هدف طراحی شده که به امر آموزش مدل سازی معادله ساختاری یاری رساند و شاید به همین دلیل است که تقریباً هیچ پیش فرضی در مدل ترسیم شده وجود ندارد و همه چیز باید توسط کاربر تعریف شود. این چنین خصیصه ای هر چند در ابتدا ممکن است در ابتدا مشکل ساز به نظر برسد اما در واقع پس از انجام چند تمرین کاربر به خوبی با انواع پارامترهای موجود در یک مدل تدوین شده آشنا می‌شود. موضوعی که در LISREL به علت پیش فرض بودن بسیاری از پارامترها اتفاق نمی‌افتد.

نخستین نسخه این نرم‌افزار در سال ۱۹۶۸ پس از تأسیس «نرم‌ن»^۱ منتشر شد، شرکت سازنده این نرم‌افزار همان شرکت سازنده نرم‌افزار spss می‌باشد که از ورژن ۱۶ به بعد نرم‌افزار spss این نرم‌افزار نیز بر روی آن موجود است. آموس فراتر از توانمندی‌های معمول نرم‌افزارهای مدل سازی رفته و به عنوان مثال به خوبی می‌تواند بر اساس جدیدترین روش‌های آماری در مورد نحوه برخورد با داده‌های مفقود شده به جایگزینی آنها دست زند. نگارش جدید آموس نه تنها کلیه ویژگی‌های نرم‌افزارهایی نظیر لیزرل را داراست بلکه خصایص منحصر به فردی دارد که آن را از سایر نرم‌افزارهای مدل سازی متمایز ساخته است.

آموس مدل ترسیم شده در صفحه میانجی را به عنوان مدل می‌پذیرد و خروجی‌های آن به خوبی و با کیفیت بالا قابل انتقال به سایر برنامه‌های تحت ویندوز است. کاربرد این امکان را دارد که با تغییر قلم، رنگ، ضخامت پیکان‌ها، اندازه پارامترها و مکان قرار گرفتن آنها مدل مدون را به زیباترین شکل و مطابق با سلیقه خود درآورد. یکی از مهم‌ترین خصایص این نرم‌افزار در ویرایش جدید آن است که با استفاده از روش بیزی می‌تواند به برآورد پارامترها در مدل‌هایی بپردازد که متغیرهای حاضر در آن‌ها از نوع مقوله‌ای رتبه‌ای یا اسمی هستند. خصیصه ای که در نگارش اولیه این نرم‌افزار وجود نداشت. به طور کلی مدل سازی معادله ساختاری بیزی از خصایص بسیار مهم و جذاب Amos است.

Sources & References:

<http://www.ssmr.ir.۱>

<http://www.iranresearches.ir.۲>

<http://analysisacademy.com.۳>

<http://www.imca.ir.۴>

راحتی اسکرپیت‌ها و برنامه‌های مورد نظر خود را در این نرم‌افزار پیاده سازی کنید. شرکت StatSoft این امکان را به شما م‌دهد تا برای شرکت خودتان نسخه خاص از نرم‌افزار که برای کار شما بهینه شده است را داشته باشید.

اگرچه در ابتدای بحث از قابلیت‌های بالای گرافیکی این نرم‌افزار صحبت شد اما روشهای تحلیل در این نرم‌افزار هم بسیار کامل بوده و خصوصاً در قسمت مدل سازی دارای منوهایی بسیار کامل است و این درحالی است که با توجه به کامل بودن منوها دسترسی و کار با منوها و پیاده سازی مدل‌ها بسیار آسان می‌باشد.

* قابلیت‌های کلیدی نرم‌افزار Statistica:

۱. تقسیم‌بندی مباحث علمی و نرم‌افزاری تحت عنوان ماژول‌های مربوط
۲. قابلیت سفارشی‌شدن، خودکار عمل کردن و توسعه‌یافتن
۳. قابلیت انجام آماره‌های توصیفی، ضریب همبستگی، محاسبه احتمالات و...
۴. قابلیت رسم و ارائه بیش از صد نوع نمایش گرافیکی با اتصال پویا بین نمودارها و داده‌ها (یعنی اتصالی که با تغییر داده‌ها نمودارها تغییر می‌کند).
۵. قابلیت برازش انواع توزیع‌ها به داده‌های عددی
۶. قابلیت انجام آنالیز واریانس و کوواریانس یک‌طرفه و چند طرفه
۷. انواع رگرسیون خصوصاً انجام رگرسیون با هزاران متغیر
۸. قابلیت انجام آنالیز داده‌های ناپارامتری
۹. قابلیت رسم نمودارهای کنترل کیفیت و تحلیل آن‌ها و نیز ویرایش این نمودارها
۱۰. قابلیت تجزیه و تحلیل سری‌های زمانی و طراحی و تحلیل آزمایشات
۱۱. قابلیت حل مسائل از طریق شبکه‌های عصبی و همچنین وجود یک سیستم خبره پیشنهاد کننده آسان و در دسترس

• نرم افزار Lisrel:

Lisrel (لیزرل) مخفف linear structural relations است. لیزرل یک محصول نرم‌افزاری است که به منظور برآورد و آزمون مدل‌های معادلات ساختاری طراحی و از سوی «شرکت بین‌المللی نرم‌افزار علمی» به بازار عرضه شده است. این نرم‌افزار با استفاده از همبستگی و کوواریانس اندازه‌گیری شده، می‌تواند مقادیر بارهای عاملی، واریانسها و خطاهای متغیرهای مکنون را برآورد یا استنباط کند و از آن می‌توان برای اجرای تحلیل عاملی اکتشافی، تحلیل عاملی مرتبه دوم، تحلیل عاملی تأییدی و همچنین تحلیل مسیر (مدلیابی علت و معلولی با متغیرهای مکنون) استفاده کرد.

یکی از قوی‌ترین و مناسب‌ترین روش‌های تجزیه و تحلیل در تحقیقات علوم رفتاری و اجتماعی، تجزیه و تحلیل چند متغیره است زیرا اینگونه موضوعات چند متغیره بوده و نمیتوان آنها را با شیوه دو متغیری (که هر بار یک متغیر مستقل با یک متغیر وابسته در نظر گرفته می‌شود) حل نمود. «تجزیه و تحلیل ساختارهای کوواریانس» یا همان «مدلیابی معادلات ساختاری»، یکی از اصلی‌ترین روش‌های تجزیه و تحلیل ساختار داده‌های پیچیده و یکی از روش‌های نو برای بررسی روابط علت و معلولی است و به معنی تجزیه و تحلیل متغیرهای مختلفی است که در یک ساختار مبتنی بر تئوری، تأثیرات همزمان متغیرها را به هم نشان می‌دهد. از طریق این روش می‌توان قابل قبول بودن مدل‌های نظری را در جامعه‌های خاص با استفاده از داده‌های همبستگی، غیر آزمایشی و آزمایشی نمود. روش لیزرل ضمن آن-که ضرایب مجهول مجموعه معادلات ساختاری خطی را برآورد می‌کند برای برازش مدل-هایی که شامل متغیرهای مکنون، خطاهای اندازه‌گیری در هر یک از متغیرهای وابسته و مستقل، علیت دو سویه، هم‌زمانی و وابستگی متقابل می‌باشد طرح ریزی

داده کاوی و پروفسور نخعی زاده

سیده صدیقه ذکی پور

تهیه کننده :



* چکیده :

امروزه داده کاوی به عنوان یکی از ابزارهای پرتوان در تحلیل داده‌ها و پایگاه های داده ای حجیم با ابعاد زیاد بین محققان در علوم مختلف از جمله محققان آمار شناخته شده است . بر همین اساس مراکز آماری مختلف روز به روز به اهمیت داده‌کاوی پی برده و این مهم را در اولویت های برنامه‌ای خود قرار داده اند . هدف این مقاله معرفی داده کاوی، شاخه های مرتبط با آن، اهمیت و کاربرد آن است .

* مقدمه :

از سال ۱۹۵۰ به بعد که رایانه در تحلیل و ذخیره سازی داده ها به کار رفت ، حجم اطلاعات ذخیره شده در آن با گذشت زمان بیشتر شده و همچنین رو به فزونی است. بسیاری از پایگاه های داده ها چنان گسترش یافته اند که شامل چند صد میلیون یا چندین میلیارد رکورد ثبت شده هستند و امکان تحلیل و استخراج اطلاعات روش های عمده و کلاسیک آماری از این پایگاه داده ها مستلزم داشتن دانش و ابزارها توان مندتر است. بنابراین نیاز به طراحی سیستم های که قادر به اکتشاف صریح اطلاعات موردعلاقه کاربران با تاکید بر حداقل مداخله انسانی باشد از یک سو و روی آوردن به روش های تحلیل مناسب با حجم داده های زیاد از سوی دیگر به خوبی احساس می شود . در حال حاضر داده کاوی معتبرترین فناوری برای بهره برداری موثر، صحیح و سریع از داده های حجیم بوده و از اهمیت بالایی برخوردار است.

* تاریخچه داده کاوی :

داده کاوی^۱ فرآیندی است که در آغاز دهه ۵۰ مطرح شد و با نگرشی نو، به مسئله استخراج اطلاعات از پایگاه داده ها^۲ می پردازد. از سال ۱۹۹۵ داده کاوی به صورت جدی وارد مباحث آمار شد و در سال ۱۹۹۶، اولین شماره مجله کشف دانش و معرفت از پایگاه داده ها منتشر شد. محققانی نظیر برآچمن و آناند (۱۹۹۶) کلیه مراحل واقع گرایانه و روبه جلو کشف دانش از پایگاه داده ها را تشخیص دادند. در حال حاضر، داده کاوی مهمترین فناوری جهت بهره برداری موثر از داده های حجیم است و اهمیت آن رو به فزونی است. به طوریکه تخمین زده شده است که مقدار داده ها در جهان هر ۲۰ ماه به حدود دو برابر می رسد.

در سال ۱۹۸۹ و ۱۹۹۱ کارگاه های کشف دانش و معرفت از پایگاه دادهها توسط پیاتسکی^۳ و همکارانش برگزار شد. در فواصل سال های ۱۹۹۱ تا ۱۹۹۴ کارگاههای کشف دانش و معرفت از پایگاه دادهها توسط فیاد^۴ و پیاتسکی و دیگران برگزار شد. به طور رسمی اصلاح داده کاوی اولین بار توسط فیاد در اولین کنفرانس بین المللی «کشف دانش و داده کاوی»^۵ در سال ۱۹۹۵ مطرح شد. امروزه کنفرانس های مختلفی در سراسر دنیا در این زمینه برگزار می شود.

* داده با اطلاعات چه فرقی دارد؟

بسیاری از مردم به اشتباه می پندارند که داده با اطلاعات تفاوت ندارد و عموماً این دو واژه را به جای هم استفاده می کنند. داده (Data) می تواند هر نوع از کاراکتر شامل متن، عدد، کلمه، صدا و تصویر باشد و در صورتی که توسط انسان مشاهده شود لزوماً معنای خاصی هم در بر نخواهد داشت. داده ها عموماً خام، دسته بندی و طبقه بندی نشده هستند و در صورتی که بخواهیم از آنها به صورت مستقیم استفاده کنیم عموماً بی فایده خواهد بود.

حال آنکه پس از طبقه بندی، دسته بندی و ساختاردهی به داده ها اطلاعات (Information) به وجود می آید. می توان از داده ها برای تصمیم گیری و یا

ایجاد دانش در مورد یک مقوله استفاده کرد. اطلاعات عموماً برای کاربر مفهوم دارد و قابل استفاده است.

مثال: تاریخ دمای سراسر جهان در صد سال گذشته داده است، حال آنکه روند افزایش یا کاهش دما در این سالها یک اطلاعات است.

نمرات دانشجویان یک کلاس داده است، اما ترتیب نمرات، میانگین، و ارتباط نمره این درس با معدل دانشجو یا یک درس دیگر یک نمونه از اطلاعات است. داده کاوی چیست؟

نگاهی به ترجمه تحت الفظی داده کاوی ، ما را در درک بهتر واژه کمک می کند. mine به معنای نهفته و با ارزش زمین اتلاق می شود. پیوند این کلمه با کلمه داده، جستجوی عمیق از داده های قابل دسترس با حجم زیاد برای یافتن اطلاعات مفید که قبلاً نهفته بود تاکید دارد. تعاریف متنوعی از داده کاوی در مراجع مختلف و توسط افراد مختلف ارائه شده است. از جمله :

۱. داده کاوی عبارت است از فرآیند استخراج اطلاعات معتبر قابل فهم و قابل اعتماد از پایگاه داده های بزرگ و استفاده از آن در تصمیم های تجاری یزرگ
۲. داده کاوی یعنی تجزیه تحلیل مجموعه داده های قابل مشاهده برای یافتن روابط مطمئن بین داده ها
۳. عبارت داده کاوی مترادف با یکی از عبارات های استخراج دانش، برداشت اطلاعات، واری اطلاعات است که در حقیقت کشف دانش در پایگاه داده ها (KDD) را توصیف می کند. اما تعریفی که در اکثر مراجع به اشتراک گذاشته شده عبارت است از استخراج اطلاعات و دانش و کشف الگو های پنهان از پایگاه داده های بسیار بزرگ و پیچیده.

* انبوه داده بلای جان سازمان ها

یک کسب و کار فرضی فعال در زمینه فروش را در نظر بگیرید، این سازمان هر روز صدها مورد فروش را از دهها مشتری ثبت می کند، تمام داده های مربوط به یک خرید از جمله نام و قیمت و دسته کالاها و اطلاعات مربوط به خریدار را ثبت می کند. پس از گذشت مدتی این سازمان انبوهی از داده ها بی معنا دارد که نمی تواند از آنها بهره برد. این سازمان اگر بخواهد بفهمد کدام مشتریان از چه کالایی بیشتر خوششان آمده راهی ندارد، نمی تواند بفهمد خریداران با هزینه بالا از چه کالایی خرید می کنند و خریداران با سبد قیمتی پایین چه کالایی را می پسندند؛ نمی داند کدام کالا فروش بهتری دارد، چه کالایی در انبار می ماند، چه کالایی حجم سرمایه سازمان را درگیر می کند و در یک کلام: از میان انبوه داده ها هیچ دانش مفیدی استخراج نمی کند. داده کاوی کمک می کند تا سازمان ها با کاوش بررسی داده های یک سیستم ، الگو و رفتارهای آینده را کشف و پیش بینی کرده و بهر تصمیم بگیرند .

داده کاوی با استفاده از تحلیل وقایع گذشته یک تحلیل اتوماتیک بیش بینانه ارائه می نماید و به سوالاتی جواب می دهد که پاسخ آنها در گذشته ممکن نبود و یا به زمان زیادی نیاز داشته است.

* ویژگی های داده کاوی :

۱. به فرضیه احتیاجی ندارد
۲. ابزارهای داده کاوی از انواع مختلف داده ها (نه تنها عددی) می توانند استفاده کنند.
۳. الگوریتم های داده کاوی به طور اتوماتیک روابط را ایجاد می کنند.
۴. داده کاوی به داده های صحیح و درست نیاز دارد.
۵. نتایج داده کاوی نسبتاً پیچیده می باشد و نیاز به متخصصانی جهت بیان آنها به

مدیران دارد.

۶. اساس داده کاوی بر مبنای سه فعالیت اصلی حذف داده های بی ارزش، فشرده سازی داده ها و کشف الگوها است.

* مراحل داده کاوی:

داده کاوی، شامل مراحل زیر است:

۱. شناخت: اولین مرحله این است که ما حوزه موضوعی و کاربری مورد نظری را که قرار است داده کاوی در آن انجام شود، بشناسیم. نیازها و هدف های کاربران را بررسی کرده، دانشی در آن باره کسب نماییم.
 ۲. داده های هدف: مجموعه هدف ها را انتخاب کنیم و داده های مورد نظر خود را انتخاب نماییم.
 ۳. پاک سازی داده ها: داده های اضافی و نامربوط از مجموعه داده ها حذف می شوند.
 ۴. تبدیل و تغییر شکل داده ها: داده های انتخاب شده به صورتی متناسب جهت انجام فرایند داده کاوی تبدیل می شوند.
 ۵. داده کاوی: این مرحله، سخت ترین مرحله است و در آن از تکنیک های هوشمند همچون: قوانین انجمنی، خوشه بندی، طبقه بندی و ... برای استخراج الگوهای مفید استفاده می شود.
 ۶. ارزیابی و تفسیر الگوها.
 ۷. دانش.
- شکل: فرایند داده کاوی

* کارهایی که در داده کاوی انجام می شوند:

شش عمل و وظیفه مهم را می توان برای داده کاوی به شرح ذیل برشمرد که سه مورد اول در گروه داده کاوی هدایت شده و دو مورد بعدی، در گروه داده کاوی غیرهدایت شده هستند:

- طبقه بندی
- تخمین
- پیش بینی
- گروه بندی شباهت
- خوشه بندی

• توصیف و نمایه سازی

* ابزارها و نرم افزارهای داده کاوی:

با توجه به اهمیت داده کاوی در دنیای کسب و کارهای بزرگ و حرفه ای، ابزارها و نرم افزارهای متعددی برای این منظور طراحی و توسعه داده شده اند. برخی از این ابزارها، رایگان و متن باز^۱ هستند و برخی دیگر به صورت بسته های نرم افزاری تجاری ارائه شده اند. از این میان، تعدادی از مواردی که بیشتر مورد استفاده هستند، در ادامه آمده است.

بدون شک، زبان برنامه نویسی و بسته نرم افزاری R یکی از مهم ترین و کارآمدترین ابزارها در زمینه تحلیل و استنتاج آماری و انجام انواع محاسبات است. زبان برنامه نویسی R، دارای امکانات فراوانی، برای انجام عملیات داده کاوی و پیاده سازی الگوریتم های مربوط به آن است. این بسته نرم افزاری، کاملا رایگان و متن باز می باشد.

یکی دیگر از ابزارهای مهم و کاربردی در زمینه داده کاوی، نرم افزار اکسل^۲ است، که به صورت پیش فرض و بعضا با افزودن برخی افزونه های تجاری، امکان انجام عملیات داده کاوی را فراهم می آورد.

نرم افزار ریپید مایتر^۳ نیز، یکی دیگر از نرم افزارهای تخصصی داده کاوی است که امکانات متعددی را برای انجام انواع عملیات داده کاوی، یادگیری ماشین، پردازش

متن، پیش بینی و تحلیل اقتصادی-مالی، تدارک دیده است. نسخه های قدیمی تر این نرم افزار به صورت متن باز منتشر شده اند؛ اما نسخه جدید آن، به صورت تجاری در دسترس قرار گرفته است.

بسته نرم افزاری Weka نیز، یکی دیگر از مجموعه های نرم افزاری متن باز است که به زبان جاوا پیاده سازی شده است، و گروهی مستقر در دانشگاه وایکاتو^۴ در کشور نیوزلند، مسئولیت توسعه و نگهداری این بسته نرم افزاری را بر عهده دارند. این مجموعه نرم افزاری به صورت اختصاصی برای انجام عملیات یادگیری ماشین پیاده سازی شده است، که طبعاً کاربردهای فراوانی را می تواند در حوزه داده کاوی داشته باشند. این بسته نرم افزاری به صورت رایگان و متن باز توزیع شده است.

نرم افزار و زبان برنامه نویسی متلب^۱، به عنوان یک نرم افزار بسیار پر کاربرد، دارای امکانات بسیار زیادی برای رشته های مختلف است، که از آن جمله می توان به موضوع تحلیل و استنتاج آماری، یادگیری ماشینی، سیستم های فازی، شبکه های عصبی مصنوعی، مدل سازی، بهینه سازی و پیش بینی اشاره کرد، که همگی دارای کاربردهای گسترده در داده کاوی هستند. در کنار قابلیت های موجود در نرم افزار متلب، می توان با استفاده از زبان برنامه نویسی متلب نیز، الگوریتم های جدید و مورد نیاز را نیز پیاده سازی کرد. هسته اصلی نرم افزار متلب به صورت تجاری توزیع شده است. اما برخی کتابخانه ها و جعبه ابزارهای رایگان نیز برای انجام عملیات داده کاوی توسط گروه های پژوهشی و دانشگاهی مختلف قابل استفاده هستند.

* کاربردهای داده کاوی:

کاربردهایی که برای داده کاوی وجود دارند، بسیار بسیار گسترده اند و ما در این نوشتار، فقط امکان معرفی تعداد محدودی از آن ها را داریم. به عنوان مثال های بیشتر، می توان به کاربردهای داده کاوی در زمینه های زیر اشاره کرد:

- سیستم های مدیریتی، مثلا مدیریت ارتباط با مشتریان
- نرم افزارهای امنیتی، مثلا نرم افزاری مانیتورینگ شبکه و ویروس کش ها
- سیستم های بانکی، مثلا تخصیص اعتبار به مشتریان و طبقه بندی آن ها
- مالی و اقتصادی، مثلا پیش بینی قیمت یک یا چند سهام یا شاخص
- پیش بینی مربوط به بازار بورس
- تحلیل سبد خرید
- شناسائی طبقات و گروه های اصلی مشتریان
- برنامه ریزی و مکان یابی، مثلا چینش داخلی فروشگاه های بزرگ و یا تخصیص امکانات شهری
- علوم پزشکی، مثلا پیش بینی خطرات احتمالی ناشی از یک عمل جراحی خاص
- علوم اجتماعی و سیاسی، مثلا پیش بینی نتایج انتخابات
- پردازش اطلاعات جمع آوری شده از فضا و اطلاعات مربوط به سفینه های فضائی
- ارائه دانش مفید برای اتخاذ تصمیم نهائی جهت پرتاب یا عدم پرتاب یک سفینه به فضا

* محدودیت‌های داده کاوی:

در حالیکه محصولات داده کاوی ابزارهای قدرتمندی می‌باشند، اما در نوع کاربردی کافی نیستند. برای کسب موفقیت، داده کاوی نیازمند تحلیل گران حرفه‌ای و متخصصان ماهری می‌باشد که بتوانند ترکیب خروجی بوجود آمده را تحلیل و تفسیر نمایند. در نتیجه محدودیت‌های داده کاوی مربوط به داده اولیه یا افراد است تا اینکه مربوط به تکنولوژی باشد.

اگرچه داده کاوی به الگوهای مشخص و روابط آن‌ها کمک می‌کند، اما برای کاربر اهمیت و ارزش این الگوها را بیان نمی‌کند. تصمیماتی از این قبیل بر عهده خود کاربر است. برای نمونه در ارزیابی صحت داده کاوی، برنامه کاربردی در تشخیص مظنونان تروریست طراحی شده که ممکن است این مدل به کمک اطلاعات موجود در مورد تروریست‌های شناخته شده، آزمایش شود. با اینهمه در حالیکه ممکن است اطلاعات شخص بطور معین دوباره تصدیق گردد، که این مورد به این منظور نیست که برنامه منظونی را که رفتارشان به‌طور خاص از مدل اصلی منحرف شده را تشخیص بدهد.

تشخیص رابطه بین رفتارها یا متغیرها یکی دیگر از محدودیت‌های داده کاوی می‌باشد که لزوماً روابط اتفاقی را تشخیص نمی‌دهد. برای مثال برنامه‌های کاربردی ممکن است الگوهای رفتاری را مشخص کند، مثل تمایل به خرید بلیط هواپیما درست قبل از حرکت که این موضوع به مشخصات درآمد، سطح تحصیلی و استفاده از اینترنت بستگی دارد. در حقیقت رفتارهای شخصی شامل شغل (نیاز به سفر در زمانی محدود) وضع خانوادگی (نیاز به مراقبت پزشکی برای مریض) یا تفریح (سود بردن از تخفیف دقیق پایانی برای دیدن مکان‌های جدید) ممکن است بر روی متغیرهای اضافه تأثیر بگذارد. در شماره‌های بعدی نشریه کارا به صورت تخصصی تر به بحث داده کاوی خواهیم پرداخت.

* پروفیسور نخعی زاده:

پروفیسور غلامرضا نخعی زاده متولد ۹ اردیبهشت ۱۳۲۷ (April ۲۹، ۱۹۴۸) در مشهد است. وی لیسانس ریاضیات خود را از دانشگاه مشهد، فوق لیسانس آمار خود را از موسسه آمار و انفورماتیک تهران و دکترای خود را از دانشگاه کارلسروهه آلمان دریافت کرد. پروفیسور نخعی زاده از سال ۱۹۹۷ تا ۲۰۰۶ به عنوان پژوهشگر برجسته مرکز تحقیقات شرکت دایملر کرایسلر آلمان^{۱۱} (مرسدس بنز) در زمینه «یادگیری ماشین و داده کاوی» مشغول به فعالیت بودند. وی مسئول بیش از ۴۰ پرسنل شرکت دایملر کرایسلر بوده است که شامل مدیران تیم‌های مختلف، محققان و دانشجویان کارشناسی ارشد و دکتری می‌شد و همگی بر روی قسمت‌های مختلفی از پروژه‌های تحقیقاتی شرکت دایملر کرایسلر کار می‌کردند.

در این مدت او مدیریت پروژه StatLog (ارزیابی الگوریتم‌های یادگیری ماشین) را نیز بر عهده داشت که با حمایت بازار مشترک اروپا و با همکاری بسیاری از دانشگاه‌های اروپایی صورت گرفت. در این مدت موضوعاتی چون شبکه‌های عصبی، یادگیری ماشین بصورت نمادین، استنباط موردی، آمار محاسباتی، تخلیص اطلاعات و ارزیابی اطلاعات، در دپارتمان داده کاوی شرکت دایملر کرایسلر بر روی تکنولوژی‌های اصلی شرکت استفاده شدند و توسعه پیدا کردند.

علاوه بر این یک پروژه پژوهشی بنیادی مرکز تحقیقات شرکت دایملر کرایسلر به نام «داده کاوی با استفاده از یادگیری ماشین» تحت ریاست پروفیسور نخعی زاده به اجرا درآمد. موضوع این پروژه، توسعه، ارزیابی و کاربرد روش‌های مختلف داده کاوی بود. در چهارچوب این پروژه راه حل‌های متعددی بر پایه داده کاوی برای کاربردهایی نظیر «مدیریت گارانتی»، «مدیریت کیفیت» و «رتبه بندی اعتباری» ارائه گردیدند.

از فعالیت‌های دیگر دپارتمان داده کاوی شرکت دایملر کرایسلر^{۱۲} در مدت ریاست پروفیسور نخعی زاده، همکاری گسترده این دپارتمان در اجرای پروژه CRISP-DM (فرآیند استانداردسازی استفاده فراصنعتی داده کاوی) بود که با حمایت بازار مشترک اروپا صورت گرفت و هدف آن تدوین یک استاندارد برای کمک به اجرای پروژه‌های داده کاوی بود.

تاکنون چندین جلد کتاب و مقالات با موضوع داده کاوی و کاربردهای آن به چاپ رسانده است. پروفیسور نخعی زاده ده‌ها کارگاه بین‌المللی داده کاوی، یادگیری ماشین و اقتصادسنجی کاربردی برگزار کرده است و علاوه بر برگزاری آن، وی به عنوان سخنران و مدرس داده کاوی و یادگیری ماشین، به دانشگاه‌هایی از کشورهای مختلف دعوت شده است؛ از جمله: اتریش، فرانسه، آلمان، ایتالیا، پرتغال، اسپانیا، سوئیس، ترکیه، انگلستان و ایالات متحده آمریکا.

پروفیسور نخعی زاده، هم‌اکنون به عنوان استاد دانشگاه کارلسروهه آلمان و همچنین مشاور ارشد در انجام پروژه‌های داده کاوی مشغول به فعالیت می‌باشد.

منابع و مراجع:

۱. خدیجه مرادی، «استخراج مفهوم در داده کاوی»
۲. دبیرخانه دائمی کنفرانس داده کاوی ایران (Permanent Secretariat of Iran Data Mining Conference)
۳. ایمان اشکاوند راد «داده کاوی»
۴. گروه داده کاوی ایران (<http://www.irdatamining.com>)

۱ Data mining	۷ Excel
۲ Knowledge discovery database(KDD)	۸ Rapid Miner
۳ Piatetsky	۹ University of Waikato
۴ Fayyad	۱۰ MATLAB
۵ Knowledge Discovery and Data Mining	۱۱ University of Karlsruhe
۶ Open Source	۱۲ DaimlerChrysler

معرفی کتاب

محمد اسمعیل زاد

تهیه کننده :



از مهم ترین اهداف ترویج علم، بسط مفاهیم علمی در سطح جامعه و آسان سازی آن برای درک همگان است. از این رو در این بخش از نشریه با هم به معرفی دو کتاب آماری خواهیم پرداخت.

• کتاب ترفندها و سواد آماری در مطالعات اقتصادی و اجتماعی

همانگونه که سواد خواندن و نوشتن بر استفاده از کلمات متمرکز شده، سواد آماری بر تصمیم‌گیری‌ها با استفاده از آمار به عنوان سند و مدرک متمرکز شده است. سواد آماری به ما کمک می‌کند تا داده‌های زندگی خود را به اطلاعاتی قابل اعتماد و درست تبدیل می‌کنیم و توسط آن برنامه‌ریزی صحیحی داشته باشیم. این کتاب با تمرکز بر روش‌های آماری معمول در مطالعات اقتصادی-اجتماعی، مقارن با بیست و پنج سالگی انجمن آمار ایران، افتتاح خانه آمار اصفهان و بزرگداشت روز جهانی آمار منتشر شده است تا گامی در ارتقای سواد و فرهنگ آماری کشور برداشته شود. مولف این کتاب آقای افشین آشفته است.

• کتاب یک روز با آمار

کتاب «یک روز با آمار» توسط آقای «افشین آشفته» در قالب یک داستان به آموزش نکاتی از سواد آماری می‌پردازد که با تایید انجمن آمار ایران و توسط انتشارات فاطمی به مناسبت روز جهانی آمار به چاپ رسیده است. در حال حاضر این کتاب در مراسمی برگزار شده به مناسبت های مختلف آمار پخش شده و تعدادی نیز در اختیار انجمن دبیران ایران قرار داده شده است.

فایل PDF هر دو کتاب از سایت انجمن علمی آمار^۱ دانشگاه گیلان قابل دانلود می باشد.





بیابید از این روز جهانے آمار تاریخے با اعلام و بزرگداشت
نقش آمار در پیشرفت جوامع خودیک موفقیت
بسازیم. (سال ۲۰۱۵)

بان کے مون دبیر کل سازمان ملل بہرہبران جهان بہ
مناسبت روز جهانے آمار

شماره اول نشریہ کارا
نشریہ برگزیدہ در جشنواره حرکت داخلے دانشگاه
گیلان (سال ۱۳۹۷)
نشریہ شایسته تقدیر در حوزه علوم پایه یازدهمین
جشنواره ملے حرکت (سال ۱۳۹۷)