علوم آماری:
تحول، محرکها و جهت‌گیری‌ها•

پیتر هال*
ترجمهٔ روح‌الله جهانی‌پور

در این نوشته، تحول علوم آماری را طی قرن گذشته به اختصار مرور می‌کنیم و توجه خاصی به عوامل و انگیزه‌های اصلی آن داریم: کاربردها، نظریه‌های ریاضی که زیربنای آن را تشکیل می‌دهند، و اخیراً، پیشرفت محاسبات الکترونیکی که به تحقق آن یاری رسانده است. خواهیم دید که تأثیرات متقابل این جنبه‌های مختلف، پیچیده است و علی‌رغم بعضی تنشهایی که ممکن است (مثلاً بین هواداران نظریه از یک سو و کاربردها از سوی دیگر) رخ دهد، مستقیماً برای آمار سودمند است. پیشرفت در ابعاد گوناگون، یقیناً جنبهٔ ممتاز آمار در میان علوم ریاضی است و البته هم منشأ قوت آن است (مثل نقشی که آمار به‌عنوان تنها ابزار ریاضی در تحلیل و تعبیر داده‌های آزمایشی و مشاهده‌ای ایفا می‌کند) و هم منشأ ضعف آن (به‌خصوص اینکه آمار هیچ جایگاه طبیعی در هیچ نظام علمی خاصی، حتی شاید در خودِ آمار هم ندارد). به همین دلیل، سیر پیشرفتهای آتی آمار تا اندازه‌ای از اختیار آماردانان هم خارج شده است. از دیدگاهی دیگر، این مرحله از مسیر تحول آمار، نشان‌دهندهٔ بازگشت به ریشه‌های علم آمار است که تنوعِ بسیار دارند. این ریشه‌ها، به‌عنوان نمونه، در نجوم، کشاورزی، انسان‌شناسی، جانورشناسی، روانشناسی، اقتصاد و ریاضی جای دارند. این مرحله، به اضافهٔ رشد خارق‌العادهٔ توان محاسباتی، حاکی از تغییرات اساسی هم در نحوهٔ کار و تفکر آماری و هم در نوع روشهایی است که در آینده به‌کار خواهند رفت. با این حال، اجتناب‌ناپذیریِ این گذار، خود حاکی از این است که بحث بر سر درستی یا نادرستی آن، چیز چندانی عاید ما نمی‌کند. به هر تقدیر، وضعیت کنونی آمار به یقین مبنای صحیحی برای پیشرفتهای آتی است، درست همان‌طور که گوناگونی سرچشمه‌ها، مبنای محکمی را در سالهای اولیهٔ رشد این موضوع برای آن فراهم کرده بود.

۱. پیشرفتهای قرن بیستمی

تئودور پورتر تاریخدانی در دانشگاه کالیفرنیا در لُس‌آنجلس، دربارهٔ آمار می‌نویسد: «اگر نگوییم هیچ‌وقت، لااقل از زمان ابداع حسابان به این طرف، هیچ شاخهٔ جدیدی از ریاضیات تا به این اندازه دامنهٔ کاربرد وسیعی نداشته است». وی در کتابش با عنوان ظهور تفکر آماری در سالهای ۱۸۲۰ تا ۱۹۰۰ که به سال ۱۹۸۶ منتشر شد [۱۹]، تمام توجه خود را به آن سالهای قرن نوزدهم معطوف نمی‌کند، بلکه در همان بند آغازی این کتاب، محرکهای متعدد و دستاوردهای علوم آماری در قرن بیستم توصیف می‌شود:

> «در قرن بیستم، آمار به‌عنوان ابزار ریاضی تحلیلِ داده‌های آزمایشی و مشاهده‌ای شناخته شده است. آمار که در سیاستگذاریهای رسمی آن را تنها مبنای مطمئن برای قضاوت دربارهٔ مثلاً تأثیر روشهای درمان یا بی‌خطر بودن مواد شیمیایی می‌دانند، و اهل صنعت و تجارت از آن در مواردی همچون کنترل کیفیت صنعتی سود می‌جویند، به وضوح در میان آن دسته از محصولات علم جای دارد که تأثیرشان بر زندگی عمومی و خصوصی افراد بسیار گسترده و فراگیر است. همچنین تحلیل آماری برای حصول نتایج قابل اعتماد از مشاهدات تجربی، در بسیاری از رشته‌های علمی اجتناب‌ناپذیر

شده است. در حوزه‌های جدیدی همانند ژنتیک کمّی، مکانیک آماری و آزمون هوش در روانشناسی، ریاضیات آماری جزء لاینفک نظریهٔ مربوط شده است.»

هرچند قرن بیستم شاهد توسعه و شکوفایی حوزه‌های بسیاری از ریاضیات بوده است، تعداد اندکی از آنها در طی عمرشان دچار دو انقلاب بزرگ شده‌اند، و آمار یکی از آنهاست. انقلاب نخست، پیدایش و نشر یک علم ریاضی جدید به منظور تحلیل داده‌هایِ متضمن خطاهای تصادفی بود. روشهای این علم آنقدر به قرن بیستم تعلق دارد که منصفانه می‌توان گفت: «پایه‌های آمار ریاضی در سالهای بین ۱۸۹۰ تا ۱۹۳۰ ریخته شد.»[۱۹]. در زمانی که پورتر کتاب خود را نوشت، یعنی تقریباً پانزده سال پیش، هنوز هم صحت این مطلب قابل دفاع بود که «روشهای اصلی تحلیل داده‌های عددی طی همین دوره پایه‌ریزی شدند».

ولی امروزه از نظر بسیاری از آماردانان این ادعا باید تعدیل شود. در همین پانزده سال، سرعت و حافظهٔ رایانه‌ها، هزار برابر شده است و انقلاب دوم در آمار که یقیناً متأثر از، و شاید هم ناشی از، پیشرفت در فن محاسبات رایانه‌ای بوده است، باعث تغییرات اساسی در روشهای آماری گشته و این تازه ابتدای راهی طولانی است و در چند دههٔ آینده، موضوع آمار به چیزی تبدیل خواهد شد که از نظر بُرد و نوع مسائل مورد بررسی و تأکید، با آنچه امروزه می‌شناسیم، تفاوت بسیار خواهد داشت. اگر پیشرفتهای قبلی آمار در بستر پیشرفتهای امروزیِ محاسبه رخ داده بود، آنگاه به احتمال زیاد این موضوع کمتر جنبهٔ ریاضی می‌داشت و بیش از آنچه اکنون هست، یک علم تجربی به شمار می‌رفت.

بخش بزرگی از آمار در انتهای قرن نوزدهم و ابتدای قرن بیستم از کاربردهایی در علوم زیستی و اجتماعی و نیز کشاورزی نشأت گرفت و تا اندازهٔ زیادی متخصصان این حوزه‌های علمی بودند که علم آمار را توسعه دادند و بخشی از نسل اول آماردانان را تربیت کردند. مثلاً فرانسیس گالتون که شاید بیشتر به خاطر تحقیقات پیشگامانه‌اش دربارهٔ هوش انسان مشهور باشد، انسان‌شناس و سیاح بود. فلورانس نایتینگل که هم در پرستاری و هم در ریاضیات آموزش دیده بود، اطلاعات زیادی دربارهٔ آمار مرگ و میر داشت و این اطلاعات انگیزه‌ای برای نوآوری و ابتکار او در جمع‌آوری، ارائه و تعبیر داده‌های آماری شد. کارل پیرسون، که ریاضیدان بود، یقیناً به خاطر روشهایی شهرت یافت که در ضمن به‌کاربردن آمار در مسائل زیست‌شناسی مربوط به وراثت و تکامل ابداع کرد. ویلیام گوست[1] (استودنت)، شیمیدانی که دستی هم در ریاضیات داشت، نخستین کارها را در زمینهٔ مسائل کوچک‌نمونه‌ای در آمار انجام داد. رافائل ولدن[2]، که جانورشناس بود، نخستین تحقیقات را دربارهٔ زیست‌سنجی انجام داد و تأسیس مجله‌ای را که بیومتریکا نامیده شد به منظور ثبت تحقیقات بیشتر در این زمینه، پیشنهاد کرد. فرانسیس اجوورث[3] زبانشناس بود و ریاضیات را هم پیش خود آموخته بود، لکن مهمترین کار او در آمار مربوط به زمانی می‌شود که به‌عنوان اقتصاددان فعالیت می‌کرد. چارلز اسپیرمن[4] که معادلِ ناپارامتری ضریب همبستگی را ابداع کرد، روانشناس بود. رانلد فیشر درجهٔ کارشناسی نجوم را از دانشگاه کیمبریج اخذ کرده بود ولی ریاضیات هم می‌دانست و انگیزه‌اش از مطالعه در آمار، علاقه‌اش به خطاهای مشاهده‌ای در داده‌های نجومی بود. با وجود این، کارهای اصلی او در آمار عمدتاً به سال ۱۹۱۹ مربوط می‌شود که با انستیتوی تحقیقات کشاورزی در روتمستد[1] همکاری می‌کرد و هم آن موقع بود که کارهای پیشگامانه‌ای در زمینهٔ طرحهای آزمایشی و تحلیل واریانس انجام داد.

البته، یک روند بسیار تأثیرگذار دیگر هم در علوم آماری در قرن بیستم وجود داشت که محرک اصلیش خودِ ریاضیات بود نه اینکه از ریاضیات به‌عنوان ابزاری صرف برای توصیف و توسعه بهره جوید. واربرگ [۲۴] به‌خصوص کارهای آبراهام والد[2] در آمار نظری را نقطهٔ شروع «دورهٔ چهارم» در گسترش آمار می‌داند. والد آموزش ریاضی بسیار پیشرفته‌ای دیده بود و علاقهٔ اصلیش نظریهٔ فضاهای متریک و هندسهٔ دیفرانسیل بود، با این حال ماندگارترین کارهای علمی او در زمینهٔ آمار، به‌خصوص تحلیلهای دنباله‌ای و نظریهٔ توابع تصمیم است.

در قسمت نظری آمار، گذشته از اینها، می‌توان گفت که پیشرفت علم احتمال در قرن بیستم و به‌خصوص آن بخشی که به مکاتب آمریکایی، فرانسوی و روسی مربوط می‌شود، تأثیر زیادی بر علم آمار داشته است. لکن مناقشات ناشی از جنگ سرد [بین شوروی سابق و غرب] از یک سو و دیوار شیشه‌ای که متأسفانه آماردانان را به نظری و تجربی تقسیم‌بندی کرده است، مانع از گسترش ارتباط بین احتمال و آمار شده است. به‌عنوان نمونه‌ای از مسألهٔ اول، یادآور می‌شویم که حتی در سال ۱۹۹۱ اندکی پس از اوجگیری گلاسنوست در شوروی، روزنامهٔ بریتانیایی ساندی‌تایمز در شمارش هزارتن از سازندگان قرن بیستم، هیچ جایگاهی برای آندری کولموگوروف قائل نشده بود در حالی‌که برخی از دانشمندان نه چندان تأثیرگذار غربی (و اتفاقاً رانلد فیشر و کارل پیرسون) را در فهرست خود آورده بود. کولموگوروف در درجهٔ اول ریاضیدانی با استعداد و قوهٔ ابداع فوق‌العاده و در درجهٔ دوم احتمال‌دان بود. طیف علائق علمیش، از منطق ریاضی و کوهومولوژی گرفته تا حرکت سیارات و جریان متلاطم هوا در موتورهای جت، گسترده بود. با وجود این، بی‌شک شهرتش بیشتر به خاطر دستاوردهای او در نظریهٔ احتمال است، تا جایی که دایرةالمعارف بریتانیکا، احتمال را «تخصص اصلی» او می‌داند. دستاورد کولموگوروف دربارهٔ ویژگیهای فرایندهای تصادفی، گذشته از کاربردهای دیگرش، ارتباط مستقیمی با نظریهٔ آماری اطلاعات و از آن طریق، با علوم رایانه‌ای دارد.

البته، هدف اصلی من از شرح کارهای شخصیتهای آمار قرن بیستم، نشان دادن گستردگیِ فوق‌العادهٔ ریشه‌های آمار است نه توصیف فعالیتهای علمی تک‌تک بنیانگذاران آن. می‌توانستیم به نام افراد دیگری هم اشاره کنیم، مثلاً: هارالد کرامر[3] که همانند گوست هم شیمیدان بود و هم ریاضیدان؛ بوریس گنه‌دِنکو[4]، هرلد جفریز[5]، جرزی نیمن[6] که همچون فیشر هم فیزیکدان بودند و هم ریاضیدان؛ پراسانتا ماهالانوبیس[7] که علاوه بر این پیش‌زمینه، امتحان تریپوس در علوم طبیعی را هم در دانشگاه کیمبریج گذرانده بود. تِد هنان[8]، که مانند اجوورث از پیش‌زمینهٔ اقتصادی خود تأثیر می‌گرفت؛

1. William Gossett 2. Raphael Weldon 3. Francis Edgeworth
4. Charles Spearman

1. Rothamsted 2. Abraham Wald 3. Harald Cramér
4. Boris Gnedenko 5. Harold Jeffreys 6. Jerzy Neyman
7. Prasanta Mahalanobis

هرلد هتلینگ[1] که هم در روزنامه‌نگاری و هم در ریاضیات آموزش دیده بود و بسیاری افراد دیگر. کمتر شاخه‌ای از علوم ریاضی در قرن بیستم بوده است که بتوان ادعا کرد افرادی با پیش‌زمینه‌هایی تا این حد متنوع، جزو مبدعان برجستهٔ آن بوده‌اند.

۲. محرکهای متعدد آمار

شرحی که در بخش قبل راجع به پیدایش روشهای امروزی آمار آوردیم، چگونگی انگیزه‌بخش بودن مسائل عملی را نشان داد. حتی کارِ والد در زمینهٔ تحلیلهای دنباله‌ای که تا اندازهٔ زیادی مرهون دقت و وسعت معلومات ریاضی او بود، از لزوم استانداردهای بالا درکنترل کیفیت صنعتی طی جنگ دوم جهانی انگیزه گرفته بود.

با وجود این، بخش اعظم جنبهٔ نظری آمار و احتمال بدون هیچ‌گونه ارتباط مستقیمی با کاربردها رشد کرده است. در موارد مهمی، نظریه، انگیزه‌بخش، نیرودهنده و تأثیرگذار بوده و خود منجر به کاربردهای خاصی شده است. برای مثال، انگیزهٔ اصلی فِلِر در تکمیل کارِ کولموگوروف در صورت‌بندی نظریهٔ احتمال به کمک نظریهٔ اندازه صرفاً ریاضی بود، لکن طی این کار به طیف وسیعی از کاربردهای احتمال به‌ویژه در ژنتیک علاقه‌مند شد. یا کارهای کولموگوروف در زمینهٔ فرایندهای مارکوفی زمان‌ـپیوسته و نیز نظریهٔ ارگودیک، آثار کاربردی خاصی داشته است (از جمله کاربرد در نظریهٔ تلاطم شاره‌ها و پدیده‌های وابسته به آن). با این حال، دیدگاههای کولموگوروف دربارهٔ کاربردهای «درست»، مانع از گسترش مرزهای نظریه می‌شد. برای مثال، مشهور است که کولموگوروف دربارهٔ کاربرد بالقوهٔ خودـنگاشت‌های دایره در نظریهٔ ضربان قلب، به یکی از شاگردانش [گویا آرنولد] توصیه کرده بود که «این از آن دست مسائل کلاسیکی نیست که شایستهٔ تحقیق باشد». مسألهٔ کاربرد در ضربان قلب همچنان ناتمام ماند تا آنکه بالاخره ۲۵ سال بعد مورد توجه مجدد قرار گرفت.

شایان ذکر است که برای مدتی طولانی، پیشرفت آمار تأثیر نسبتاً کمی از پیشرفت محاسبات الکترونیکی می‌گرفت. محاسبات رایانه‌ای که از نظر تاریخی شاید بتوان گفت از میانهٔ دههٔ ۱۹۴۰ آغاز شده، در دست‌کم یک ربع قرن نقش چندان مهمی در آمار ایفا نمی‌کرد. با این حال، استثناهای قابل‌توجهی هم وجود داشت، از جمله کارِ جان توکی همراه با کولی در مورد الگوریتم تبدیل فوریهٔ سریع [۴]، یا کلیتر از آن، تشخیص زودهنگام توکی در مورد تأثیر عظیمی که رایانه‌های رقمی می‌توانستند داشته باشند، و نیز توسعهٔ محاسبات آماری در روتمستد در بریتانیا. اما به رایانه‌ها، چه دستی و چه الکترونیکی، تا اندازهٔ زیادی، به چشم ابزارهایی برای کشف ویژگیهای روشهای موجود نگاه می‌شد تا به‌عنوان وسیله‌ای برای ایجاد روشهای دیگر. البته، توانِ نسبتاً کم ماشینهای محاسبهٔ قدیمی هم از جملهٔ دلایل چنین وضعی بود.

برای مثال، یکی ازکارهای اولیه در «مطالعهٔ شبیه‌سازی» را اگون پیرسون، پسر کارل، انجام داد و به کمک آن ویژگیهای گوناگون ضرایب همبستگی رتبه‌ای متعارف را روشن ساخت. [رجوع کنید به [۷]، [۸]، [۱۸]]. روش مونت‌کارلو که عمدتاً در فیزیک مطرح شد و برای محاسبهٔ انتگرالها و کمیتهای وابسته به آنها به‌کار رفت، تا دههٔ ۱۹۷۰ نقش مهمی در آمار بازی نمی‌کرد، تا آنکه نخستین کارها را در این زمینه جان همرزلی[1] و ریاضیدانان دیگری انجام دادند، هرچند ممکن است آماردانان آنها را از خودشان بدانند. البته، استثناهایی هم وجود داشت مثل کاری که جورج بارنارد [۱] در مورد آزمون مونت‌کارلو انجام داد و همین‌طور تحقیقاتی که روی روشهای نمونه‌گیریِ مجدد صورت گرفت که بعداً دربارهٔ آنها توضیح خواهیم داد.

درواقع، یکی از دلایل اصلی [عدم رویکرد به ماشینهای محاسبه] این بود که اکثریت آماردانان تا حدود نیمهٔ دوم دههٔ ۷۰ به آسانی به رایانه‌های الکترونیکی پرقدرت دسترسی نداشتند. البته، توکی خیلی پیشتر یعنی از سال ۱۹۴۵ با آزمایشگاههای بل همکاری می‌کرد و از آن طریق به رایانه دسترسی داشت. اما تا دو یا سه دههٔ آخر قرن بیستم، برای اکثریت آماردانان، محاسبات الکترونیکی یا ناممکن بود و یا در بهترین حالت، فرایندی پردردسر و غیرتعاملی.

اکنون، در آغاز قرن جدید، وضعیت کاملاً تغییر کرده است. طی دو دههٔ گذشته، محاسبات رایانه‌ای همراه با کاربردهای بالقوه و نظریه‌های ریاضی به‌عنوان یکی از محرکان اصلی و نه فقط یک عامل تسهیل‌کننده، در ایجاد و تکامل روشهای آماری مؤثر بوده است. محاسبات رایانه‌ای را دیگر به راستی می‌توان محرک و مولد دانست، زیرا در علم نوین آمار، بسیاری از مسائل ابتدا به شکل «مجرد» ــ یعنی به‌صورت مسائلی که بالقوه کاربردی‌اند اما بدون آنکه کاربرد خاصی از آنها در نظر باشد ــ بازبینی و سپس راه‌حل‌هایی برمبنای روشهای آماریِ رایانه‌ـمحور برای آنها ارائه می‌شود. نمونهٔ خوبی از این دست، ارائهٔ روشهای خودگردان برای ساختن بازه‌های اطمینان و آزمونهای فرض است (برای ملاحظهٔ جزئیات بیشتر به بخش ۳ رجوع کنید). فنون جدید و نظریه‌های ریاضیِ زیربنای آنها، عمدتاً به واسطهٔ وجود ابزارهای محاسباتی که می‌توان به وقت نیاز از آنها استفاده نمود، توسعهٔ فراوان یافته‌اند. سالها بود که مبانی احتمالاتی برای پیشبرد مرزهای نظریه وضع شده بود، لکن بدون دسترسی به ابزارهای محاسباتیِ موردنیاز، انگیزه و توجیه کافی برای پرداختن به مسائل فراهم نبود.

بحث فوق گویای این است که ارّابهٔ محاسبات، از جنبه‌های مهمی، در جلو اسبِ انگیزه‌های کاربردی بسته شده است. در هیچ کجای تاریخ گذشتهٔ آمار، محاسبات (که امروزه برمبنای توان رایانه‌های الکترونیکی سنجیده می‌شود) هیچ‌گاه در پیشبرد روشهای آمار چنین محرک مهمی که اکنون هست نبوده است. البته، بعضیها از این وضع متأسف‌اند و نظیر همان دغدغه‌هایی را دارند که دربارهٔ نقش محرّک نظریه‌های ریاضی در پیشبرد آمار ابراز می‌شود.

ولی این دغدغه‌ها دربارهٔ نقش نظریه و محاسبه، نابه‌جا هستند زیرا برمبنای دیدگاهی نسبت به تحول آمار استوارند که همهٔ ابعاد آن را در نظر نمی‌گیرد، دیدگاهی که انگیزه‌ها و محرکهای گوناگون را در یک طیف خطی ایده‌آل کنار هم قرار می‌دهد یا در بهترین حالت آنها را از طریق یک نمودار مسطح به هم مربوط می‌سازد. در این دیدگاه، تصور بر این است که خدمت‌رسانی علم آمار به جامعه، فرهنگ و نیز صنعت در صورتی بهینه خواهد بود که نتایج عملی در وهلهٔ نخست موردنظر قرارگیرد و سپس تحقیقات

1. Harold Hotelling

1. John Hammersley

قهقرایی برای رسیدن به مبانی انجام شود. اما در حقیقت، ارتباطهای بهینه خیلی پیچیده‌ترند و در راستای هر مسیر ارتباطی، انگیزشهایی وجود دارد.

به خصوص، قبلاً به اهمیتی اشاره کردیم که انگیزه‌های ریاضی برای عالمان آمار، نظیر فِلِر و کولموگوروف داشته است. آمار نظری و محاسبهٔ رایانه‌ای توانسته‌اند با روشهایی که گاهی توصیف دقیق آنها مشکل است ولی حقیقتِ وجودی دارند، از طریق ارائهٔ مضامین مشترکی که فراتر از تقسیم‌بندیهاست، وحدتی به آمار ببخشند. یگانه‌سازی حتی از خود آمار هم فراتر رفته است. برای مثال، دقیقاً ۳۰ سال پیش، کسانی که از دیدگاه ریاضی در مهندسی سیستمها نظریه‌پردازی می‌کردند، روشهای خود گردانِ خاصی را برای براورد پارامترهای آماری ارائه دادند (مثلاً رجوع کنید به [۲۱]، [۲۰]). این روشها، با خودگردان افرون ارتباط دوری دارند، از این نظر که در آنها از روشهای تکرار عددی برای ساختن براوردکننده‌هایی استفاده می‌شود که عملکرد بهتری دارند، مثلاً براوردکننده‌هایی که اُریبی کمتری دارند. از آن طرف در بستر محاسبات هم تعداد زیادی از دست‌اندرکارانِ حوزه‌های متعدد، با نرم‌افزارهای آماری از قبیل S^+ و R آشنا (و حتی متکی به آنها) شده‌اند و لذا رهیافتهایشان به حل مسائل با رهیافت آماردانانی که آن نرم‌افزارها را ارائه داده‌اند، یکسان شده است.

تحقق و توضیح بسیاری از مباحث جاری و مهم در آمار معاصر، نیازمند نظریه‌های ریاضی کاملاً پیشرفته‌ای بوده است. روشهای موجکی یکی از این موارد است. تعریف و ساختن بسطهای موجکی تجربی و حتی درک درست مزیتهای عملی روشهای متفاوت، نیازمند به‌کارگیری آنالیز ریاضی جدید است. به همین صورت، بعضی از مفاهیم اساسی در نظریهٔ تحلیل تابعی داده‌ها، مانند مفهوم چگالی روی فضای توابع و روشهای استنباط آماری دربارهٔ ساختار آن، محتاج ابزارها و مفاهیم پیچیده‌ای از ریاضیات است. حاصل به‌وجود آمدن روشهایی از این دست این است که امروزه ارتباط بین نظریه و کاربردها، خیلی مهمتر و پیچیده‌تر از آن است که قبلاً بود.

این پیچیدگی، تا اندازه‌ای از این واقعیت نشأت می‌گیرد که بسیاری از روشهای جدید آماری، نتایجی به دست نمی‌دهند که بتوان آنها را به‌طور صریح از روی داده‌ها حساب کرد، بلکه فقط می‌توان، مثلاً، به کمک روشِ مونت‌کارلو، تقریبهایی عددی از آنها به‌دست آورد. فرایندهای آماری جدید و سازوار[1]ی (مثل فرایند داده‌کاوی) وجود دارند که در آنها کوشش بر این است که نوعی مصالحه بین پیچیدگی نتیجه و سادگی تعبیر آن برقرار شود. به همین صورت، نظریهٔ حاکم بر پیشرفتهایی از این دست، نیازمند نوعی انتزاع ریاضی است که از یک طرف آنقدر انعطاف‌پذیر باشد که جنبه‌های اساسی آماری مسألهٔ در دستِ بررسی را حفظ کند، و از طرف دیگر آنقدر نیرومند باشد که بتواند نتایج نظری را به‌طور شفاف، به خصوص برای کاربران آمار بیان کند. از این‌رو، در حالی‌که صورت‌بندی روشهای آماری ارتباط تنگاتنگی با روشهای محاسباتی رایانه‌ای پیدا می‌کند، استدلالهای ریاضی موردنیاز برای توصیف این روشها، بیش از پیش محتاج مهارتهای آماردانان نظری می‌شود. درواقع طنز قضیه در این است که، توجه زیاد علوم آماری جدید به محاسبات رایانه‌ای به جای فرمولهای سادهٔ جبری، نیاز به ریاضیات خلاق و حرفه‌ای را بیشتر کرده است.

1. adaptive

ضمناً، بخشی از قویترین حمایتها از پیشرفت محاسبات رایانه‌ای و الگوریتمهای وابسته به آن، در سالهای اولیه، از جانب دانشمندان علوم ریاضی به عمل آمده است، نه لزوماً از جانب دانشمندان دیگری که از محاسبات الکترونیکی بیشترین فایده را درکارهایشان می‌برده‌اند. برای مثال، فرانک ییتس[1]، آماردان مشهورکه از دانشگاه کیمبریج مدرک ریاضی گرفته بود، نقش اصلی را در تأسیس انجمن رایانهٔ بریتانیا داشت و مدتی هم رئیس آنجا بود. او در جایی می‌نویسد: «برای اینکه کسی آماردان نظری خوبی باشد باید محاسِب خوبی هم باشد و لذا باید بهترین ابزارهای محاسباتی را هم در اختیار داشته باشد». البته، حمایت ریاضیدانان از محاسبات رایانه‌ای گاهی هم با مخالفتهای شدیدی روبه‌رو شده است. نمونه‌اش کولموگوروف است که علی‌رغم مخالفتهای شدیدی که در اتحاد شوری با او می‌شد، از ریاضیات محاسباتی و نظریهٔ محاسبات دفاع می‌کرد. نتیجه‌اش این شد که بسیاری از دانشمندان علم رایانه در شوروی سابق از دههٔ ۱۹۶۰ تا ۱۹۸۰ شاگردان یا شاگردانِ شاگردان کولموگوروف بودند.

۳. روشهای نمونه‌گیری مجدد: یک مطالعهٔ موردی

ارائهٔ روشهای نمونه‌گیری مجدد یا روشهای خودگردان مثال روشنگری است از اینکه چگونه جنبه‌های متفاوتی از علم، از قبیل کاربردهای خاص، ریاضیات، و محاسبهٔ رایانه‌ای، می‌توانند در زمانهای متفاوت طوری گرد هم آیند که محرک و ایجادکننده مجموعهٔ واحدی از روشها باشد. مناسبترین تعریف امروزی روشهای نمونه‌گیری مجدد این است که در این روشها در تعریف یک کمیت مجهول مثل یک پارامتر، به جای توزیع آماری مجهول، همتای تجربی آن را قرار می‌دهند. به این ترتیب، میانگین نمونه، براوردکنندهٔ خودگردانِ میانگین جامعه خواهد بود و این نشان می‌دهد که ریشه‌های اصلی روشهای خودگردان را باید در روزگاران گذشته جُست.

اما، در نظرگرفتن رابطهٔ بین نمونهٔ مجددِ به‌تصادف انتخاب شده با نمونهٔ اصلی به‌عنوان «مدلی» برای رابطهٔ نمونه و جامعه‌ای که نمونه از آن انتخاب می‌شود، ایدهٔ نوینی است و می‌توان گفت به اواسط قرن بیستم برمی‌گردد. به خصوص پراسانتا ماهالانوبیس [۱۷]، ریاضیدان هندی، از این رویکرد استفاده کرد تا اطلاعاتی دربارهٔ تغییرپذیری آزمونهای میدانی به‌دست آورد. امروزه، این روش را «روش خودگردان انتخاب m شیء از میان n شیء بدون جایگذاری» می‌نامیم. کار ماهالانوبیس انگیزهٔ عملی خاصی داشت: تحلیل داده‌های مربوط به محصول برنج در فصل برداشت ۱۹۳۴ـ۱۹۴۴، در ایالات بیهار هند. کار لیلیان و ویلیام مَدو [۱۶] و نیز هاوارد جونز [۱۲] هم، به روشهای خودگردان امروزی ارتباط دارد، هرچند این روشها مستقیماً از آن نتیجه نشده‌اند. مک‌کارتی ([۱۴]، [۱۵]) نیز از طراحان اولیهٔ زیرنمونه‌گیری تصادفی بود، منتها این بار در زمینهٔ بررسی نمونه‌ای.

جان هارتیگان ([۱۰]، [۱۱]) این ایده را بسیار گسترش داد، منتها بیشتر از جنبهٔ روش‌شناسی تا از نظر کاربردهایی که ممکن بود داشته باشد. نظریهٔ او دقت و توضیح لازم ــ و در حد خودش، انگیزه و توجیه اساسی ــ برای این فن فراهم کرد. در سوی دیگر طیف، جولیان سایمن جامعه‌شناس (که اقتصاددانی معتقد به بازار آزاد، لیبرال، و فردی جامع‌الاطراف بود)، طرحی

1. Frank Yates

برای روش‌شناسی آمار برمبنای شبیه‌سازی مونت‌کارلو ارائه داد که بسیار شهودی و در عین حال آینده نگرانه بود (رجوع کنید به [۲۲]، بخش ۴). طرح سایمن عاری از هرگونه نظریهٔ ریاضی، و بنابراین عاری از بینشها و مفاهیمی اساسی بود که می‌شد از آن طریق آنها را به‌دست آورد و انتقال داد. اتکای این طرح بر شبیه‌سازیهای تصادفی تکراری، در زمانی‌که محاسبهٔ فشردهٔ نتایج آماری به یک اصل تبدیل شده بود و رایانه‌های پرقدرتی که به آسانی در دسترس باشند هنوز دوران طفولیت خود را می‌گذراندند، باید عجیب و غریب به نظر آمده باشد، لکن این طرح سایمن، چندان بیشتر از طرحهای دیگرش با مخالفت روبه‌رو نشد (پس از مرگ سایمن در سال ۱۹۹۸، بن واتنبرگ[1] در یادداشتی به مناسبت فوت او که در نشریهٔ وال‌ستریت چاپ شد از او به‌عنوان سرمایه‌داری دو آتشه یاد می‌کند که «با طرفداران محیط زیست و مالتوسی‌ها درگیر مبارزهٔ فکری شد»). با این حال، رویکرد سایمن منعکس‌کنندهٔ دیدگاه تجربی درستی دربارهٔ علوم اجتماعی بود.

کار سایمن قدیمیتر از آن است که متأثر از امکانات محاسباتیِ سهل‌الوصول بوده باشد، و همان‌طور که گفتیم مسلماً از نظریه هم الهام نگرفته است. درواقع، براد افرون [۵ و ۶] بود که ارتباط حیاتی بین دو جنبهٔ محاسباتی و نظریِ کارهای سایمن را هوشمندانه و به طرزی بسیار مؤثر برقرار کرد. افرون مشاهده کرد که وقتی به جای یک توزیع نمونه‌گیری واقعی، براوردکننده‌اش را قرار می‌دهیم، مفهوم «پارامتر» را وسیعتر از آنچه قبلاً تصور می‌شد می‌توان تعبیر کرد. درواقع، این تعبیر از پارامتر، پارامترهای تنظیم‌کننده در انواع زیادی از فرایندهای براورد و خیلی چیزهای دیگر را شامل می‌شود (حتی ترازهای اسمیِ بازه‌های اطمینان و آزمونهای فرض را می‌توان پارامترهای تنظیم‌کننده تلقی کرد). سپس این ایده‌ها را به منابع محاسباتی نیز ربط داد: هنگامی‌که توابع توزیع نمونه‌گیری را نمی‌توان مستقیماً حساب کرد، در صورت وجودِ امکان محاسباتی کافی، می‌توان آن را با دقت دلخواه به روش مونت‌کارلو تقریب زد. افرون، روشهای خودگردان را به سمتی سوق داد که می‌بایست در دو دههٔ بعد از آن دنبال شود و به‌علاوه نام این روشها را هم او بر آنها نهاد.

پس از آن، توان محاسباتی و نظریهٔ ریاضی مشترکاً عامل ایجاد و تکامل روشهای خودگردان بوده دامنهٔ کاربردهای بالقوهٔ آنها را وسعت داده‌اند. مثلاً، نظریهٔ ریاضی تقریب توزیعی، نقش مهمی در رفع ابهاماتی اساسی که دربارهٔ مناسبترین روشهای استفاده از خودگردان وجود داشت، بازی کرده است. هرچند این نظریه از کارهای چبیشف و اجوورث در حدود یک قرن پیش سرچشمه می‌گیرد، لکن تا اندازهٔ زیادی بر پیشرفتهای بسیار جدیدتر احتمال و آمار نظری استوار است، به‌ویژه آن بخشی که رابی باتاچاریا و ر. رنگه رائو توصیف کردند [۲]. این نظریهٔ ریاضی و نظریه‌های پیش از آن، ارتباط چندان مستقیمی با پیشرفتهای جاری در روشهای آماری نداشته‌اند، با وجود این، دیدگاهی به‌دست می‌دهند که برای حصول اطمینان از دقت و صحت کاربردهای روشهای خودگردان در حوزه‌های دیگر علمی از پزشکی گرفته تا مهندسی، بسیار مهم بوده است. از سوی دیگر، تحولات سریع در سخت‌افزارهای محاسباتی طی دههٔ ۱۹۸۰ انگیزه‌ای شد برای پیشرفتهای کلی در روشهای خودگردانِ دوگانه، و غلبه بر آن نوع مسائلی که در نتیجهٔ مطالعات نظری به‌وجود آمده بود. این پیشرفتها هم اغلب بدون اینکه کاربرد خاصی محرک آنها باشد، حاصل شدند، هرچند بعدها، کاربردهای مهمی به‌دست آمده است.

با این پیشرفتها، نمونه‌گیری مجدد به‌عنوان روشی برای حل انواع مسائل آماری و حل بهتر مسائل سنتی تثبیت شده است. با این حال، همهٔ این روشها وابسته به این ایده است که توزیع تجربی اولیه چیز مقدسی است. عملکرد بسیاری از روشهای آماری را می‌شود با تعدیل داده‌ها به‌طور قابل‌ملاحظه‌ای بهبود بخشید و این یکی از مسیرهای پیشرفت را در آینده تشکیل خواهد داد. روشهایی عینی برای تغییر نمونه ارائه خواهد شد، از جمله: تجدید وزن داده‌ها، یا واقعاً تغییر مقدار آنها، یا افزودن «شبه داده‌ها»یی که به کمک نمونهٔ اصلی تولید شده‌اند، همین‌طور تقلیل اریبی و واریانس، یا ساختن براوردکننده‌های استوارتر، یا سازگارکردن داده‌ها با یک «الگوی کیفی» (تا مثلاً پاسخی یکنوا یا تک مُدی به یک هم‌متغیر بدهند). همین ایده‌ها را می‌شود برای دستیابی به برازش بهتری به یک الگوی پارامتری، یا تولیدِ (مثلاً) اثری از شکل توزیع پاسخ که ممکن است در نتیجهٔ فقدان داده‌های «دورافتاده» ایجاد شود، به‌کار برد. این داده‌ها به‌طور ضمنی و به کمک الگوریتمهای تنظیم‌کنندهٔ داده‌ها معلوم می‌شوند و نه به‌طور صریح توسط خود آزمایش‌کننده. این روشها، به شدت رایانه‌ـمحور خواهند بود، مخصوصاً اگر در مورد داده‌های وابسته به‌کار روند، ولی خوب! این موضوع مانعی تعیین‌کننده نیست. نمونه‌گیری مجدد با توزیع تعدیل شده یا تنظیم‌شده و تجربی سروکار دارد نه با خود توزیع اصلی.

۴. آمار به کجا خواهد رفت و چه کسی آن را به آنجا خواهد بُرد؟

بخش اول را با اشاره به پیش‌زمینه‌های بسیار متفاوتِ دانشمندانی آغاز کردیم که موضوع آمار را در انتهای قرن نوزدهم توسعه داده، آن را به قرن بیستم رساندند. در میان آنها، انسان‌شناس (مثل گالتون)، زیست‌شناس (مثل ولدُن)، اقتصاددان (مثل اجوورث)، روانشناس (مثل اسپیرمن) و ریاضیدان (مثل فیشر و کولموگوروف) وجود داشت. این مبدعان اولیه، همراه با هم، پایه‌های علمی را ریختند که طی چند دههٔ گذشته، در سطح وسیعی توسط «آماردانان» به‌کار گرفته و تدریس شده است.

با این حال، همان‌طور که بخش مهمی از آنالیز ریاضی به‌صورت شاخه‌های متعددی درآمد که بر آنالیز تکیه دارند و آن را به‌کار می‌برند (همانند ریاضیات زیستی، ریاضیات مهندسی، ریاضیات مالی و بیمه)، همین اتفاق هم برای آمار افتاد. ایده‌های مربوط به آمار بیزی، منطق فازی و روش درستنمایی ماکسیمم فیشر، کاربردهای متنوعی در مهندسی، مثلاً در دستگاههای پیشرفتهٔ اندازه‌گیری ماتریسی در دوربینها، دارند. تعداد اندکی از کاربردها را افرادی ارائه داده‌اند که در دانشکدهٔ آمار تحصیل کرده‌اند، لکن این مطلب از ارزش روشها و نتایج آنها نمی‌کاهد. داده‌کاوی، که به نظر می‌رسد با علوم رایانه و فناوری اطلاعات بیش از آمار مرتبط باشد، نمونهٔ بارزتری است. این مبحث از جنبه‌های متعدد به فنونی که آماردانان ارائه داده‌اند متکی است (مثل رده‌بندی و درختهای رگرسیون؛ برای آگاهی بیشتر به [۳] رجوع کنید.)، ولی مورد استفادهٔ وسیع افرادی است که در فناوری اطلاعات کار می‌کنند.

1. Ben Wattenberg

بعضی از مباحث آماری دیگری که در انتهای قرن بیستم پیشرفت سریعی داشته‌اند، عبارت‌اند از ژنتیک آماری (به خصوص در ارتباط با ژنوم انسانی)، یادگیری ماشین و محاسبات عصبی (بخش عمدهٔ این موضوع اساساً آماری است هرچند بیشتر ابداع‌کنندگان آن، آماردان نبوده‌اند)، و روشهای تحلیل مجموعه‌های عظیم داده‌ها. توسعهٔ روشهای آماریِ رایانه‌ـمحور با رشدی سریعتر از مبانی نظری و آزمایشهای عددی، آمار را به طرز چشمگیری در راستای جهتهای بیان شده در فوق به پیش خواهد راند.

در حال حاضر، فقط در آستانهٔ دستیابی به آن چیزهایی هستیم که می‌توان به کمک ماشینهای محاسب، به چنگ آورد. قانون مور[1] حاکی از اینکه سرعت و حافظهٔ سخت‌افزار رایانه‌ها تقریباً هر هیجده ماه دو برابر می‌شود، بیش از ۳۰ سال است که برقرار بوده و تصور هم نمی‌رود که ناگهان نقض شود. البته اگر هم این‌طور شود، قوانین جدیدی که با فناوریهای جدید محاسباتی هماهنگی داشته باشد، جای آن را خواهند گرفت. به این ترتیب، همهٔ حوزه‌های علم بالاجبار به سمت گشایش افقهای جدید و وسیع در بررسی و تعبیر داده‌ها، حرکت می‌کنند. تعداد نسبتاً کمی از این پیشرفتها در بُعد کیفیِ داده‌ها (مثلاً در تراز «نوفه») و بیشتر آنها در کمیت داده‌ها منعکس خواهد شد و لذا روشهای جدید آماری لزوماً بخش مهم و اساسی این پیشرفتها را تشکیل خواهند داد. با وجود این، مسیرهای پیشرفت، آنقدر متعدد و پیشرفتها در هر مسیر آنقدر زیاد خواهد بود که تعداد کافی آماردان برای حضور در بطن هریک از این مسیرها وجود نخواهد داشت، حتی اگر چنین فرصتِ حضوری برای آماردانان فراهم باشد که البته این‌طور نیست.

حتی معلوم نیست که آماردانان تمام اصول زیربناییِ روشهای آینده را وضع کرده و حاضر داشته باشند. جدا از فنون جدید آماری که برای بهره‌برداری از رایانه‌های سریع و سریعتر شونده ارائه خواهد شد، نغمه‌هایی از بیرون آمار هم به گوش می‌رسد که مبانی آمار نیز محتاج بازنگری‌اند (برای مثال، توضیحی دربارهٔ دیدگاه دیوید کرانتس[2] را در ادامهٔ بحث ملاحظه کنید). اگر بخواهیم فنی‌تر صحبت کنیم، روشهایی که در حال حاضر به‌کار می‌بریم، هنوز تا حد زیادی، تقریباً خطی‌اند و از جنبه‌های مهمی که البته در آینده تغییر خواهند کرد، متأثر از کاربردهای جدیدی هستند که نوع جدیدی از مسائل را پیش رو قرار می‌دهد. به نظر می‌رسد در آینده، روشهای آماری جدید فراوانی ــ بیش از پیش به‌دست کسانی که آمار کلاسیک را نیاموخته‌اند ــ ارائه خواهد شد.

این وضعیت، باعث دلواپسی‌های فراوانی در میان آماردانان در انتهای قرن بیستم شده است. مثلاً دیوید فینی در مقاله‌ای که در خبرنامه انجمن سلطنتی آمار چاپ شد [۹]، می‌نویسد:

> «ممکن است آماردانان حرفه‌ای کنترل خود را بر آنچه به نام رشتهٔ آنها انجام می‌شود ــ و حتی نگرانی خود را در این باره ــ از دست بدهند [این مقاله را] با این پرسش آغاز کردم که [آمار] به کجا می‌رود؟ و پاسخ من این است که «رو به حضیض»! و راه بازگشت طولانی و پرپیچ و خم است، ولی اگر بخواهیم به آنچه غولهای آمار ۷۵ سال پیش ارائه دادند وفادار بمانیم، چاره‌ای بجز یافتن راه بازگشت نداریم.»

بی‌شک، علم آمار در حال بازگشت به پیش‌زمینه‌های متعددی است که از آنها سرچشمه گرفته است. آمار را در آینده، همانند یک قرن پیش، کاربران آن به‌طرز چشمگیری توسعه داده و به‌کار خواهند بُرد. این بازگشت به ریشه‌ها، به شکلی کاملاً متفاوت و همراه با روشها و انگیزه‌های بسیار پربارتر و متنوعتر از آنچه در ابتدای قرن بیستم بوده، صورت خواهد پذیرفت.

فینی دربارهٔ خطرهای ملازم با این دورهٔ گذار را از جمله خطر بالقوهٔ ارتکاب خطاهای جدی، تذکار می‌دهد.این خطرها را با این تذکرات لفظی نمی‌توان کاهش داد زیرا واقعاً جدی‌اند. اما خود از ثمرات موفقیت آمارند، ثمرهٔ آن موقعیتی هستند که [۱۹] چنین توصیف می‌کند: «آمار در میان آن دسته از محصولات علم جای دارد که تأثیرشان در زندگی عمومی و خصوصی افراد بسیار گسترده و فراگیر است». اگر این مشکلات و مسائل وجود نداشتند، آنگاه موضوعی به گستردگی و تنوع آمار که امروزه مستقیماً در زندگی همهٔ افراد بشر دخیل است، نه تنها متحول و شکوفا نمی‌شد، بلکه یا مرده بود و یا در حال مرگ بود. انقلاب دوم در علم آمار به سرعت در حال شتاب‌گیری است و نباید انتظار داشت دقیقاً در همان جهاتی پیش رود که به قول فینی «غولهای آمار ۷۵ سال پیش ارائه دادند» و همانها بود که به انقلاب اول منجر شده بود.

نباید تصور کنیم که خود ما همیشه در بهترین موقعیت برای توسعهٔ رشتهٔ خود یا حتی کسب بصیرت عمیقتری دربارهٔ آن هستیم. گاهی هم نگاه از بیرون می‌تواند مفید و آموزنده باشد. از زمان اسپیرمن تا حالا، روانشناسان، روشهای آماری متعددی را برای رفع حاجات خود ارائه داده و به‌کار برده‌اند. دیوید کرانتس [۱۳] دیدگاه روانشناسان را دربارهٔ نامناسب بودن آزمونهای فرض سنّتی آماری، در بعضی زمینه‌های مهم، بیان کرده است. او قویاً معتقد است که «منازعات کنونی در روانشناسی حکایت از نقصان جدی در مبانی آمار دارد». به‌علاوه، نمونه‌هایی از روشهایی در دست است که آماردانان دانسته یا ندانسته آنها را از حوزه‌های علمی دیگر قرض گرفته‌اند. لذا، جریان روش‌پردازی آماری همیشه در آن جهتهایی که علمای قدیم آمار پیش روی ما قرار داده‌اند پیش نمی‌رود. مثلاً، اشتغال آماردانان به ارائه روشهای تصویرپردازی، گذرگاهی دوطرفه از ایده‌ها بین آمار و علوم دیگر را پدید آورده که برای هر دو طرف سودمند است.

در آغاز قرن بیستم، وقتی ادموند ویتاکر و جورج واتسن توابع خاص و نظریهٔ مربوط به آنها را به نام «آنالیز جدید» معرفی کردند [۲۵] و این موضوع در همان اوایل قرن به حوزهٔ مهندسی و فیزیک راه پیدا کرد، ریاضیدانان محض قاعدتاً نگران از دست دادن کنترل خود [بر این موضوع] شده و احساس خطر کرده‌اند. آن خطرها ممکن است به بزرگی خطرهایی که در وضعیت گذار کنونی در آمار مطرح است، نبوده باشند زیرا آنالیز ریاضی تجویزی‌تر [دارای ضابطه‌های مشخصتر] است و فلسفهٔ چندگانه‌ای بر آن حاکم نیست که کاربرد و آموزش آن نیاز به کنترل و نظارت داشته باشد. ولی مهندسان آن زمان همان‌قدر در برابر ایمنی بشر و حفظ محیط زیست مسؤول بوده‌اند که همکاران امروزی آنها که در ژنتیک (یکی از نمونه‌های مورد اشارهٔ فینی) یا طراحی راکتورهای هسته‌ای بی‌خطر (که استفاده از نظریهٔ

1. Morre 2. David Krantz

قابلیت اعتماد در آن با انتقاد آماردانان روبه‌رو شده) کار می‌کنند. نتیجه در مورد ریاضیات این شد که به جامعه، فرهنگ و صنعت سود رسید و هم از این طریق ریاضیات غنی‌تر شد، در سطح وسیعتری شناخته شد، و به‌طور کلی از سلامت بیشتری برخوردار شد.

مسلماً وضع آمار تشابه کاملی با ریاضیات ندارد؛ با وجود این، آنقدر مشابهت هست که بتوان هم نسبت به حرفهٔ آمار و هم نسبت به خود آماردانان خوشبین بود. تنوع روبه گسترش و قابل ملاحظهٔ کاربردهای آمار و تأثیرات متقابل فوق‌العاده پربازی که مابین ایده‌های آماری، روشهای محاسباتی و نظریه‌های ریاضی برقرار شده است، اهمیت و سرزندگی آمار را در آینده تضمین می‌کند.

مراجع

1. Barnard. G. A. (1963): Contribution to discussion. *J. Roy. Statist. Soc. Ser.* B **25**, 294
2. Bhattacharya, R. N. and Ranga Rao, R. (1976): *Normal Approximation and Asymptotic Expansions*. Wiley, New York
3. Breiman. L.. Friedman. J. H., Olshen, R. A., Stone, C. J. (1984): *Classification and Regression Trees*. Wadsworth, Belmont, California
4. Cooley. J. W. and Tukey, J. W. (1965): An algorithm for the machine calculation of complex Fourier series, *Math. Comp.* **19**, 297-301
5. Efron, B. (1979 a): Bootstrap methods: another look at the jackknife. *Ann. Statist.* **7**, 1-26
6. Efron, B. (1979 b): Computers and the theory of statistics: thinking the unthinkable. *SIAM Rev.* **21**, 460-480
7. Fieller, E. C., Hartley, H. O. and Pearson, E. S. (1957): Tests for rank correlation coefficients, I. *Biometrika* **44**, 470-81
8. Fieller, E. C. and Pearson, E. S. (1961): Tests for rank correlation coefficients. II. *Biometrika* **48**, 29-40
9. Finney, D. (2000): Whither... goes now the practice of statistical science? *Roy. Statist. Soc. News* **27** No. 7, 1-2
10. Hartigan, J. A. (1969): Using subsample values as typical values. *J. Amer. Statist. Assoc.* **64**, 1303-1317
11. Hartigan, J. A. (1971): Error analysis by replaced samples. *J. Roy. Statist. Soc. Ser. B* **33**, 98-110
12. Jones, H. L. (1956): Investigating the properties of a sample mean by employing random subsample means. *J. Amer. Statist. Assoc.* **51**, 54-83
13. Krantz, D. H. (1999): The null hypothesis controversy in psychology. *J. Amer. Statist. Assoc.* **94**, 1372-1381
14. McCarthy, P. J. (1966): Replication (an approach to the analysis of data from complex surveys): *National Center for Health Statistics*
15. McCarthy, P. J. (1969): Pseudo-replication: half-samples. *Rev. Int. Statist. Inst.* **37**. 239-264
16. Madow, W. G., and Madow, L. H. (1944): On the theory of systematic sampling, I. *Ann. Math. Statist.* **15**, 1-24
17. Mahalanobis, P. C. (1946): Report on the Bihar crop survey: Rabi season 1943-1944. *Sankhyā* **7**, 269-280
18. Pearson, E. S. and Snow, B. A. S. (1962): Tests for rank correlation coefficients. III. Distribution of the transformed Kendall coefficient. *Biometrika* **49**, 185-191
19. Porter, T. (1986): *The Rise of Statistical Thinking, 1820-1900*. Princeton University Press, Princenton
20. Prasad, R. M., Sinha, A. K., and Mahalanabis, A. K. (1977): Two- stage bootstrap algorithms for parameter estimation. *Internat. J. Systems Sci.* **8**, 1365-1374
21. Rowe, I. H. (1970): A bootstrap method for the statistical estimation of model parameters. *Internat. J. Control* **12**, 721-738
22. Simon, J. L. *Basic Research Methods in Social Science*. Random House, New York
23. Sunday Times Magazine, The (1991): *1000 Makers of the Twentieth Century*. Times Newspapers Ltd, London
24. Varberg, D. (1994): The development of modern statistics. In: *From Five Fingers to Infinity. A Journey Through the History of Mathematics*. Ed. F. J. Swetz. Open Court, Chicago
25. Whittaker, E. T. and Watson, G. N. (1902): *A Course of Modern Analysis*. Cambridge University Press, Cambridge, UK

* * * * * *

• Peter Hall, "Statstical science: Evolution, motivation and direction", in *Mathematics Unlimited 2001 and Beyond*, B. Engquist and W. Schmid (eds.), springer (2001).

* پیتر هال، دانشگاه ملی استرالیا