سیزده راه نگرش به

ضریب همبستگی*

جوزف لی راجرز، و. الن نایسواندر*

ترجمهٔ علی عمیدی

در ۱۸۸۵، سرفرانسیس گالتن برای اولین بار اصطلاح "رگرسیون" را تعریف، و نظریهٔ همبستگی دومتغیره را کامل کرد. یك دهه بعد، کارل پیرسن، شاخص r پیرسن را که هنوز هم برای اندازه‌گیری همبستگی به‌کار می‌رود، ارائه‌داد. مقالهٔ ما برای بزرگداشت صدمین سال اولین بحث گالتن از رگرسیون و همبستگی نوشته شده است. مقاله را با تاریخچهٔ کوتاهی آغاز می‌کنیم و سپس ۱۳ فرمول مختلف را می‌آوریم که تعریفهای محاسباتی و مفهومی متفاوتی از rاند. هر فرمول راهی برای اندیشیدن دربارهٔ این شاخص، از دیدگاه جبری، هندسی، و مثلثاتی ارائه می‌کند. نشان می‌دهیم که r پیرسن (یا تابعهای سادهٔ r) را می‌توان به صورتهای مختلف، نظیر نوعی خاص از میانگین، نوعی خاص از واریانس، نسبت دومیانگین، نسبت دوواریانس؛ شیب یك‌خط، کسینوس یك زاویه، و مماس بریك بیضی در نظر گرفت، و همچنین نشان می‌دهیم که ممکن‌است از دیدگاههای جالب دیگری نیز به‌آن نگریست.

مقدمه

ما این روزها در اواسط "دههٔ بعد از صد سالگی" همبستگی و رگرسیون هستیم. دستاوردهای تجربی و نظریی که موجب شد رگرسیون و همبستگی به صورت مطالب آماری تعریف شوند در ۱۸۸۵ به‌وسیلهٔ سرفرانسیس گالتن ارائه شدند. سپس، کارل پیرسن در ۱۸۹۵، r پیرسن را انتشار داد. این مقاله که دربارهٔ ضریب همبستگی پیرسن گفتگو می‌کند، هم زیربنا وهم چند استنباط از r را مورد بحث قرار می‌دهد که برای مدرسین آمار مفید است.

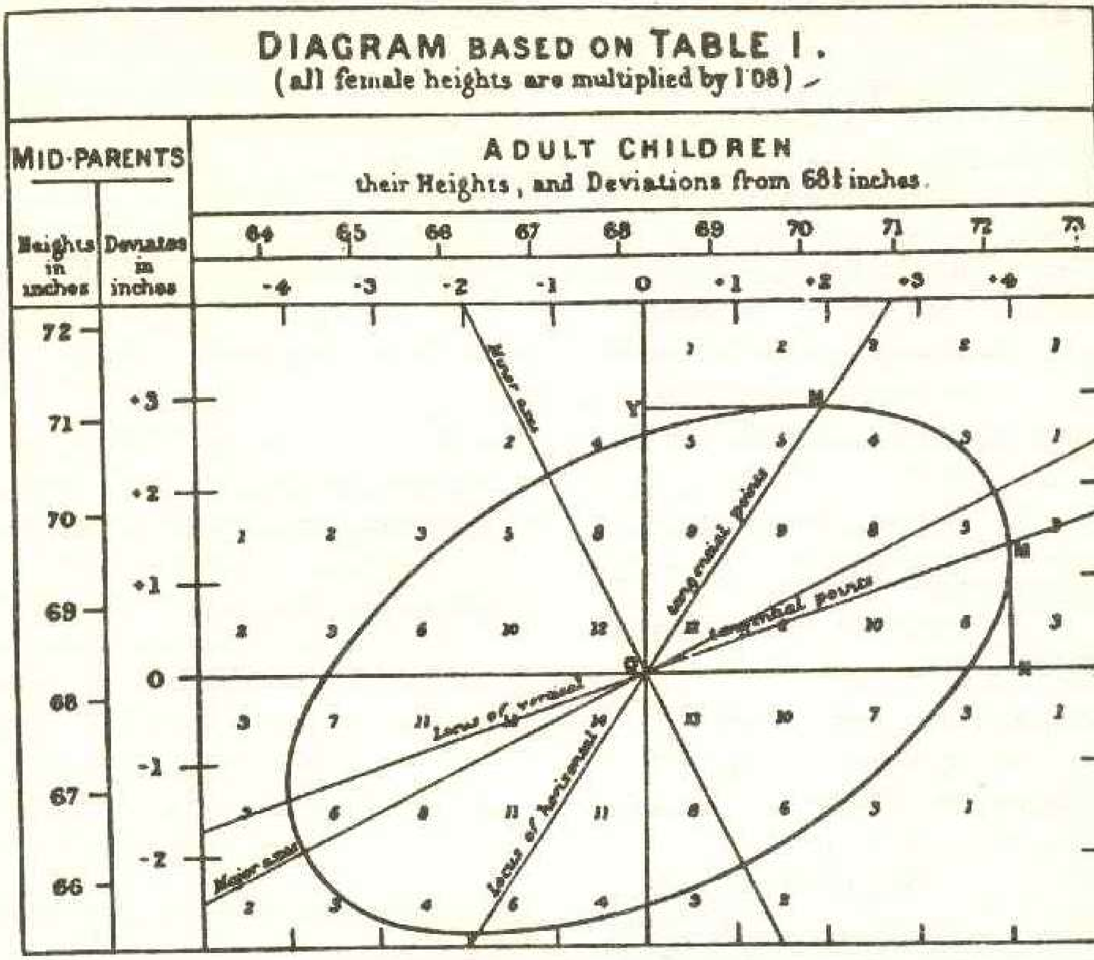
ما مقاله‌را با تاریخچه‌ای کوتاه از گسترش همبستگی و رگرسیون آغاز و به دنبال آن، راههای تعبیر ضریب همبستگی را با تفصیل بیشتر مرور می‌کنیم. این‌مرور نشان می‌دهد که همبستگی به‌صورت شاخصی کلی درآمده‌است که استنباطهای گوناگونی از آن می‌شود. مع‌هذا، گذشت زمان براین شاخص صد ساله چندان تأثیری نداشته است.

ایدهٔ اصلی همبستگی اساساً قبل از ۱۸۸۵ عنوان‌شده بود [۱۳]. پیرسن درسال ۱۹۲۰ ارائهٔ رویهٔ نرمال p متغیر همبسته (در ۱۸۲۳) را به گاوس نسبت داد. اما گاوس به‌همبستگی، به‌عنوان یك مفهوم متمایز، توجه‌خاصی نداشت و درمعادله‌های توزیعی‌اش، همبستگی را به‌عنوان یکی از پارامترها تعبیر کرد. پیرسن دریك‌مقالهٔ تاریخی قبلی که در ۱۸۹۵ انتشار یافت، ارائهٔ توزیع نرمال دومتغیره (در ۱۸۴۶) را به اوگوست براوه، ستاره‌شناس فرانسوی، نسبت داد (ر.ك. [۳۰]). براوه درواقع، به‌پارامتری از توزیع نرمال دومتغیره، عنوان " همبستگی " را اطلاق کرد، اما نظیر گاوس اهمیت‌همبستگی را به‌عنوان معیار پیوند بین‌متغیرها تشخیص نداد. [قبل‌از ۱۹۲۰، پیرسن امتیازی راکه به‌براوه داده‌بود پس گرفت. اما والکر [۴۶] و سیل[1] [۳۴] تاریخچه‌ای راکه پیرسن گزارش‌داده و درپروراندن آن کمك کرده بود مرور کردند و ادعای براوه را در تقدم تاریخی تأیید نمودند.] چارلز داروین، دایی‌زاده گالتن، در ۱۸۶۸ با اشاره به اینکه "تمام اجزاء یك نظام به میزانی معین به‌هم مربوط یا همبسته‌اند" مفهوم‌همبستگی‌را به‌کار برد. سپس، در ۱۸۷۷، گالتن دریك سخنرانی مربوط به‌بستگی مشخصه‌های جسمانی نسلهایی از والدین وفرزندان، برای‌اولین‌بار به‌اصطلاح "Reversion" اشاره کرد. "قانون Reversion" اولین توصیف رسمی از مفهومی است که گالتن بعدها "رگرسیون" را به‌جای‌آن باب کرد.

در طی این دوره، پیشرفت مهم فلسفه نیز سبب افزایش بار معنایی‌مفاهیم‌همبستگی و رگرسیون شد. در ۱۸۴۳، جان‌استوارت میل فیلسوف بریتانیایی‌برای اولین‌بار "پنج قانون تحقیق‌تجربی" خود را ارائه داد. بین‌اینها روش تغییرات ملازم هم را گنجانیده بود: "هروقت تغییر پدیده‌ای به‌روشی‌خاص، به‌نحوی‌موجب تغییر

---

1. Seal



شکل ۱. اولین نمودار پراکنش دو متغیره (از گالتن ۱۸۸۵)

پدیدهٔ دیگری شود، پدیدهٔ اول علت یا معلول پدیدهٔ دوم است، یا از طریق واقعیتی علّی به‌آن مربوط است.» میل برای معتبر بودن استباط علّی سه‌شرط پیشنهاد کرد [۶]. اولاً، علت باید از نظر زمانی مقدم بر معلول باشد. ثانیاً، علت و معلول باید به‌هم مربوط باشند. ثالثاً، توضیحات به‌ظاهر موجه دیگر باید رد شوند. بنابراین تفکیک‌پذیری همبستگی و علیت، و توصیف اولی به‌صورت شرطی لازم و نه‌کافی برای دومی، در نظام جاافتادهٔ فلسفه و نظام کم‌سابقهٔ زیست‌سنجی تقریباً به‌طور همزمان تشخیص داده شد.

تا ۱۸۸۵، زمینه برای ارائهٔ چند اثر مهم فراهم آمد. درطول آن سال، گالتن رئیس بخش مردم‌شناسی انجمن بریتانیا بود. وی درخطابه‌ای که به‌مناسبت انتصابش ایراد نمود، اولین بار به‌رگرسیون عنوان تعمیم "قانون Reversion" را اطلاق کرد. کمی بعد درهمان سال [۹]، خطابهٔ مزبور را به‌همراه اولین نمودار پراکنش دومتغیره که همبستگی را نشان می‌داد منتشر کرد (شکل ۱).

وی در این نمودار، فراوانی ترکیبهای بلندی قد فرزندان و بلندی قد والدین را به‌نمایش گذاشت. وقتی وی نتایج را هموار کرد و خطوطی بر نقاط با فراوانی برابر گذرانید، دریافت که "خطوط مار بر درایه‌های هممقدار، یک‌سری از بیضیهای هممرکز و مشابه تشکیل می‌دهند." این اولین نمایش تجربی خمهای تلث‌چنگالی از توزیع نرمال دومتغیره بود [فراوانی روی هریک از این خمها یکی است]. گالتن، با همیاری هامیلتن دیکسن، ریاضیدانی از کمبریج، توانست فرمولی نظری برای توزیع نرمال دومتغیره به دست آورد. این فرمول، از نظر ریاضی، به موضوعی که گاوس و براوه نیم قرن قبل روی آن کار کرده بودند رسمیت داد. پیرسن [۲۰، ص۳۷] اظهار داشت که "در ۱۸۸۵، گالتن نظریهٔ همبستگی دومتغیره را کامل کرد."

در سالهای بعداز ۱۸۸۵، چند پیشامد دیگر براهمیت ریاضی اثر ۱۸۸۵ گالتن افزودند. در ۱۸۸۸، گالتن متذکر شد که r، دقت "همبستگی" را اندازه می‌گیرد و (هرچند ایدهٔ همبستگی منفی هنوز به‌ذهن او نرسیده بود) اظهار کرد که r نمی‌تواند از ۱ بیشتر باشد. هفت‌سال بعد، پیرسن فرمول ریاضیی را که هنوز متداولترین فرمول برای اندازه‌گیری همبستگی است، یعنی ضریب همبستگی گشتاور حاصلضربی پیرسن را ارائه داد. از دیدگاه تاریخی، به‌نظر مناسبتر می‌رسد که نام این شاخص معروف، r گالتن-پیرسن باشد. تحولهای مهم در حکایت همبستگی و رگرسیون در جدول ۱ خلاصه شده‌اند.

اینک، پس از یک قرن، دانشمندان معاصر اغلب ضریب همبستگی را مطلبی بدیهی و مسلم می‌دانند. به این مطلب توجه نمی‌شود که قبل از گالتن و پیرسن، تنها وسیلهٔ استقرار بستگی بین متغیرها، یافتن ارتباط علّی بود. حتی راهی برای بحث دربارهٔ پیوند بین متغیرهایی که فاقد بستگی علت و معلولی بودند وجود نداشت، چه رسد به

جدول ۱. پیشامدهای برجستهٔ تاریخچه همبستگی و رگرسیون

تاریخ	شخص	پیشامد
۱۸۲۳	کارل فریدریش گاوس، ریاضیدان آلمانی	رویهٔ نرمال N متغیر تصادفی همبسته را ارائه داد.
۱۸۴۳	جان استوارت میل، فیلسوف بریتانیایی	پنج قـــانون استقرا، از جمله تغییرات ملازم را مطرح کرد.
۱۸۴۶	اوگوست بـراوه، افسر نیروی دریایی فرانسه و ستاره‌شناس	بـــا اشاره به "یك همبستگی"، روی توزیع نرمـــال دومتغیره کار کرد.
۱۸۶۸	چارلز داروین دایی‌زادهٔ گالتن، طبیعیدان بریتانیایی	تمام اجزاء یك نظام... بههم مربوط یا همبسته‌اند.
۱۸۷۷	سرفرانسیس گالتن، بریتانیایی، اولین متخصص زیست‌سنجی	برای اولین بار از "Reversion"، سلف رگرسیون، بحث کرد.
۱۸۸۵	سرفرانسیس گالتن	برای اولین بار به "رگرسیون" اشاره کرد. نمودار پراکنش دومتغیره با خمهای تك‌چگالی نرمال دومتغیره، اولین نمودار همبستگی، را انتشار داد. نظریهٔ همبستگی نرمال دومتغیره را کامل کرد[۳۵].
۱۸۸۸	سرفرانسیس گالتن	r را به‌طور مفهومی تعریف کرد، و کران بالایی آن را مشخص نمود.
۱۸۹۵	کارل پیرسن، آماردان بریتانیایی	ضریب همبستگی گشتاور حاصلضربی (گالتن-) پیرسن را تعریف کرد.
۱۹۲۰	کارل پیرسن	"یادداشتهایی بر تاریخچهٔ همبستگی" را نوشت.
۱۹۸۵		صد سالگی رگرسیون و همبستگی

اندازه‌گیری آنها. امروزه، ضریب همبستگی ــ و معادلهٔ رگرسیون همتای آن ــ در بسیاری از زمینه‌ها برای آزمایشهای مبتنی بر مشاهدات، یك ابزار آماری اصلی است. کارول [۳، ص ۳۴۷] در خطابه‌ای که به‌مناسبت انتصابش به ریاست انجمن روانسنجی ایراد کرد، ضریب همبستگی را "یکی از متداولترین ابزارهایی که در روانسنجی به‌کار می‌رود،... و شاید یکی از متداولترین ابزارهایی که بد به‌کار گرفته شده است" خواند. در تجزیهٔ عاملی، در مدلهای ژنتیکی رفتاری، درمدلهای معادله‌هـــای ساختاری (مثلاً LISREL)، و در دیگر روشهای وابسته، ضریب همبستگی به عنوان واحد اساسی داده‌هـــا به‌کار می‌رود.

بحث مـــا حول ضریب همبستگی گشتاور حاصلضربی پیرسن دور می‌زند. r پیرسن اولین معیار رسمی اندازه‌گیری همبستگی بود، و هنوز هم‌معیاری برای بستگی است که کاربرد فراوان دارد. بدون شك، بسیاری از شاخصهای همبستگی "رقیب"، در واقع حالتهای خاصی از فرمول پیرسن‌اند. ρ_Sی اسپیرمن، همبستگی نقطه‌ای مربوط به‌دوسری از داده‌ها، و ضریب فی، مثالهایی هستند که هریك با به‌کار بردن r پیرسن درمورد انواع خاصی از داده‌ها قابل محاسبه است (مثلاً، ر.ك. [۱۰]).

مطلب را با کمی چاشنی آموزشی عرضه می کنیم. درنظر اول، معیار همبستگی، ساده و سرراست است. امـــا، اختلافهای جزئی شگفت‌انگیزی درمفهوم ضریب همبستگی وجود دارند که برخی از آنها را معرفی خواهیم کرد. به‌پیروی از پیرسن، تأکید ما بر ضریب همبستگی به‌عنوان یك شاخص محاسباتی است که برای اندازه‌گیری پیوند دومتغیره به‌کار می‌رود. از نظر آماری، درك عمیقتری از همبستگی مستلزم توجه به مدل نمونه‌گیریی است که فرض می‌شود زیربنای مشاهدات باشد (مثلاً، ر.ك. [۳] و [۱۴])، و نیز مستلزم درك تعمیم همبستگی به همبستگی چندگانه و جزئی است، اما در اینجا تأکید مـــا برهمبستگی جنبهٔ بنیادیتری دارد. اولاً، تـــوجه اصلی را به وضعیتهای دومتغیره محدود می کنیم. ثانیاً، اکثر تعبیرهای ما آزاد-توزیع‌اند، زیرا محاسبهٔ همبستگی نمونه‌ای نیاز به‌هیچ فرضی دربارهٔ جامعه ندارد (ر.ك. [۱۶]). برای بررسی مسائل زیادی که به کاربرد استنباطی r (مثلاً، محدود کردن، و کوچك کردن آن) مربوط‌اند خواننده را به روشهای دیگری ارجاع می‌دهیم (مثلاً، ر.ك. [۱۲]). برای استنباط این اصلیترین معیار بستگی دومتغیره سیزده راه مختلف معرفی می کنیم. ادعا نمی کنیم که این مقاله تمام تعبیرهای ممکن ضریب همبستگی را شامل می‌شود. محققاً تعبیرهای دیگری هم وجود دارنـــد، و یقیناً تعبیرهـــای جدیدی نیز پیشنهاد خواهند شد.

۱. همبستگی به‌صورت تابعی از اندازه‌های خام و میانگینها

ضریب همبستگی گشتاور حاصلضربی پیرسن، شاخصی بی‌بعد است که تحت تبدیلهای خطی هر یك از دومتغیر ناورداست. اولین بار

پیرسن در ۱۸۹۵، فرمولی ریاضی برای این معیار مهم ارائه داد:

$$r=\frac{\Sigma(X_i-\bar{X})(Y_i-\bar{Y})}{[\Sigma(X_i-\bar{X})^2\Sigma(Y_i-\bar{Y})^2]^{1/2}} \qquad (۱.۱)$$

این فرمول، یـا صورتهای سادهٔ جبری دیگر آن، فرمولهای متداولی در کتابهای درسی آمار مقدماتی هستند. درصورت این کسر، اندازه‌های خام هرمتغیر، به‌وسیلهٔ کم کردن میانگین متغیر از آن، "مرکزی" شده‌اند ومجموع حاصلضربهای متقاطع متغیرهای مرکزی شده به‌دست آمده است. مخرج کسر، مقیاس متغیرها را برای اینکه واحدهای یکسان داشته‌باشند تعدیل می‌کند. بنا براین معادلهٔ (۱.۱)، r را به‌صورت مجموع حاصلضربهای متقاطع دومتغیر مرکزی شده و استاندارد شده توصیف می‌کند. با استفاده از نابرابری کوشی-شوارتس، می‌توان نشان داد که قدرمطلق صورت از مخرج بیشتر نیست (مثلاً، [۱۲، ص ۸۷] را ببینید). بنا براین برای r، حدود ± 1 به‌دست می‌آیند. برای منظورهـای محاسباتی، چندین تبدیل سادهٔ جبری این فرمول را می‌توان به‌کار برد.

۲. همبستگی به‌صورت کوواریانس استاندارد شده

کوواریـانس، شبیه همبستگی، معیار پیوند خطی بین متغیرهاست. کوواریانس روی مجموع حاصلضربهای متقاطع متغیرهای مرکزی شده، که مقیاس آنها تعدیل نیافته است، تعریف می‌شود. هرچند در کتابهای درسی مقدماتی کوواریانس را اغلب نـادیده می‌گیرند، ولی واریانس (که ازآن گفتگو می‌شود)، واقعاً حالتی خاص از کوواریانس است ـ بدین‌معنا کـه واریانس، کوواریانس یك‌متغیر با خودآن متغیر است. کوواریانس دومتغیر، درجامعه‌های نامتناهی، دارای کرانهای مشخصی نیست، ودر نمونه کرانهایی نامعین (و تعبیری نامناسب) دارد. بنا براین، کوواریانس اغلب‌معیار توصیفی مفیدی برای پیوندنیست، زیرا مقدارآن به‌مقیاسهای اندازه‌گیری X و Y بستگی دارد. ضریب همبستگی، ازتغییرمقیاس کوواریانس به‌دست می‌آید:

$$r=\frac{s_{XY}}{s_X\cdot s_Y} \qquad (۱.۲)$$

که درآن s_{XY}، کوواریانس‌نمونه‌ای است و s_X و s_Y انحراف معیارهای نمونه‌ای هستند. وقتی کوواریانس به دو انحراف معیار تقسیم‌شود، بردکوواریانس به بازهٔ (۱+ ,۱−) محدود می‌شود. لذا، تعبیرهمبستگی به عنوان معیار بستگی معمولاً ساده‌تر از تعبیر کوواریانس به‌عنوان‌معیاربستگی است (وبه‌وسیلهٔ آن همبستگیهای مختلف با سهولت بیشتر مقایسه می‌شوند).

۳. همبستگی به‌صورت شیب استاندارد شدهٔ خط رگرسیون

بستگی همبستگی و رگرسیون را می‌تـوان به‌صورتی ساده‌تر با رابطهٔ

$$r=b_{Y.X}\left(\frac{s_X}{s_Y}\right)=b_{X.Y}\left(\frac{s_Y}{s_X}\right) \qquad (۱.۳)$$

نمایش‌داد که درآن، $b_{Y.X}$ و $b_{X.Y}$ شیبهای خطهای رگرسیون‌اند کـه بـه‌ترتیب Y را از روی X یا X را از روی Y پیشگویی می‌کنند. در اینجا همبستگی به‌صورت تابعی‌ازشیب یکی ازدوخط رگرسیون و انحراف‌معیارهای‌دومتغیر بیان‌می‌شود. نسبت انحراف معیارها دارای این اثر است‌که واحد شیب رگرسیون را به‌واحد همبستگی تبدیل می‌کند. بنا براین همبستگی، یـك شیب استاندارد شده است.

تعبیری مشابه، همبستگی را به عنوان شیب خـط رگرسیون استانداردشده مطرح‌می‌کند. وقتی‌دومتغیر خام‌را استاندارد می‌کنیم، انحراف معیارها برابر واحد می‌شوند و شیب خط رگرسیون بـه صورت همبستگی درمی‌آید. دراین حالت، عرض ازمبدأ خط صفر است، و خط رگرسیون به‌آسانی به‌صورت

$$\hat{z}_Y = r z_X \qquad (۲.۳)$$

بیان می‌شود. از این‌تعبیر آشکار است‌که همبستگی، برای‌پیشگویی واحدهای متغیر استاندارد شدهٔ Y، واحدهای متغیر استاندارد شدهٔ X را ازنو مقیاس‌بندی می‌کند. توجه‌کنیدکه شیب رگرسیون z_Y روی z_X، خط رگرسیون را ملزم می‌کندکه درناحیهٔ سایه‌خوردهٔ شکل ۲، بین نیمسازهای محورهـای مختصات بیفتد. همبستگیهای مثبت ایجاب مـی‌کنند کـه خط از ربعهای اول و سوم بگذرد؛ همبستگیهای منفی ایجاب می‌کنندکه خط از ربعهای دوم و چهارم عبورکنند. زاویهٔ خط رگرسیون z_X روی z_Y با محور Y، برابر زاویهٔ خط رگرسیون z_Y روی z_X با محور X است، و این‌خط همان‌طورکه نشان داده‌ایم درناحیهٔ سایه‌نخوردهٔ شکل ۲ می‌افتد.

۴. همبستگی به‌صورت میانگین هندسی دوشیب رگرسیون

همبستگی‌را ممکن‌است به‌صورت تابعی همزمان از دوشیب خطهای رگرسیون استاندارد شده، $b_{Y.X}$ و $b_{X.Y}$، بیان‌کرد. درواقـع این تابع، میانگین هندسی است، و اولین‌تعبیر ازچندتعبیر r است که آن را به‌عنوان نوعی خاص ازمیانگین بیان می‌کند:

$$r=\pm\sqrt{b_{Y.X}\,b_{X.Y}}\,. \qquad (۱.۴)$$

این رابطه را می‌توان ازمعادله‌های (۱.۳) نتیجه‌گرفت: جمله‌های دوم و سوم برابریها را درهم ضرب می‌کنیم تا r^2 به‌دست آیـد، آنگاه انحراف معیارهـا را حـذف می‌کنیم، و از دوطرف جذر می‌گیریم.

تعمیمی از این‌تعبیروجود داردکه متضمن‌رگرسیون چندمتغیره است. وقتی $B_{Y.X}$ و $B_{X.Y}$، ماتریسهای‌ضرایب رگرسیون‌مربوط به‌دومجموعه‌ازمتغیرها،معلوم‌اند،جذرهای‌ویژه‌مقدارهای‌حاصلضرب این‌ماتریسها، همبستگیهای‌کانونیك این‌دومجموعه ازمتغیرها هستند. وقتی یك تك‌متغیر X و یك تك‌متغیر Y وجود داشته باشند، ایـن مقادیر به‌ضریب همبستگی ساده تبدیل می‌شوند.

۵. همبستگی به‌صورت جذر نسبت دوواریانس

گاه ازهمبستگی، به‌این دلیل‌که مقدارش تعبیرروشنی ندارد، انتقاد می‌شود. این انتقاد بـا مربع کردن همبستگی و با تعبیر زیرتخفیف می‌یابد. شاخص‌مربع‌شده را اغلب ضریب تعیین‌می‌خوانند. مقدار این ضریب را ممکن است بـه عنوان کسری از واریانس یکی از متغیرها تعبیرکرد ([۱۸] را،برای آگاهی ازبحثی‌که مربوط به‌چند

تعبیر از ضریب تعیین است، ببینید). مجموع مربعات کل ($SS_{\text{کل}}$) برای Y را ممکن است بــه دومجموع مربعات حــاصل از رگــرسیون ($SS_{\text{رگرسیون}}$) ومجموع مربعات ناشی از خطا ($SS_{\text{خطا}}$) افراز کرد. نسبتی از کل تغییرات Y که از تغییرات X ناشی می‌شود، نسبت $SS_{\text{رگرسیون}}$ به $SS_{\text{کل}}$ است، و r، جذر این نسبت است:

$$r=\sqrt{\frac{\Sigma(Y_i-\hat{Y}_i)^۲}{\Sigma(Y_i-\bar{Y})^۲}}=\sqrt{\frac{SS_{\text{رگرسیون}}}{SS_{\text{کل}}}}$$

هم‌ارز آن، اگرصورت ومخرج این‌معادله‌را به $(N-۱)^{۱/۲}$ تقسیم کنیم، r مساوی باجذر نسبت واریانسها (یــا نســبت انحراف معیارها)ی متغیرهای پیش‌بینی شده و مشاهده شده می‌شود:

$$r=\sqrt{\frac{s_{\hat{Y}}^۲}{s_Y^۲}}=\frac{s_{\hat{Y}}}{s_Y} \qquad (۱\cdot۵)$$

(توجه کنید که $s_{\hat{Y}}$، براوردی اریب از $\sigma_{\hat{Y}}^۲$ است، درحالی که $s_Y^۲$ براوردی نااریب است.) این تعبیر، تعبیری است که انگیزهٔ استنباط اولیهٔ پیرسن از شاخص همبستگی بــوده است. ([۱۵، ص ۴] را ببینید.) همبستگی به‌عنوان نسبت دوواریانس را می‌توان بــا تعبیر دیگری ازهمبستگی به‌صورت نسبت دومیانگین (منسوب به گالتن) مقایسه کرد. ما این تعبیررا دربخش ۱۳ ارائه خواهیم داد.

۶. همبستگی به‌صورت میانگین حاصلضرب متقاطع متغیرهای استاندارد شده

راه دیگر تعبیرهمبستگی به‌صورت میانگین (بخش ۴ را ببینید)، بیان آن به‌صورت متوسط حاصلضرب متقاطع متغیرهای استاندارد شده است:

$$r=\frac{\Sigma z_X z_Y}{N} \qquad (۱\cdot۶)$$

معادلهٔ (۱۰۶) را می‌توان مستقیماً ازتقسیم صورت و مخرج معادلهٔ (۱۰۱) برحاصلضرب انحراف معیارهــای دونمونه به‌دست آورد. چون‌میانگین یك‌توزیع، اولین گشتاور آن است، این‌فرمول، به‌کار بردن مفهوم "گشتاور حاصلضرب" به جــای ضریب همبستگی را توجیه می‌کند.

دو توصیف بعدی، متضمن تعبیرهای مثلثاتی از همبستگی هستند.

۷. همبستگی به‌صورت تابعی از زاویهٔ بین دوخط رگرسیون استاندارد شده

همان‌طور کــه دربخش ۳ اشاره شد، دوخط رگرسیون استاندارد شده نسبت به هرنیمساز متقارن‌اند. فرض کنیم زاویهٔ بین دوخط β باشد (شکل ۲ را ببینید). دراین‌صورت

$$r=\sec(\beta)\pm\tan(\beta). \qquad (۱\cdot۷)$$

برهان ساده‌ای از این بستگی را در اختیار داریم. معادلهٔ (۱۰۷) به‌طور شهودی واضح نیست و برای‌هدفهای محاسباتی یا مفهومی هم به‌اندازهٔ دیگریها مفیدنیست. مقدار r دراین‌رابطه نشان‌می‌دهد که بین همبستگی و فاصلهٔ زاویه‌ای دو خط رگرسیون یك بستگی متقارن وجود دارد. تعبیر بعدی، که باز تعبیری مثلثاتی است، اساساً ارزش مفهومی بیشتری دارد.

۸. همبستگی به‌صورت تابعی از زاویهٔ بین دوبردار متغیر

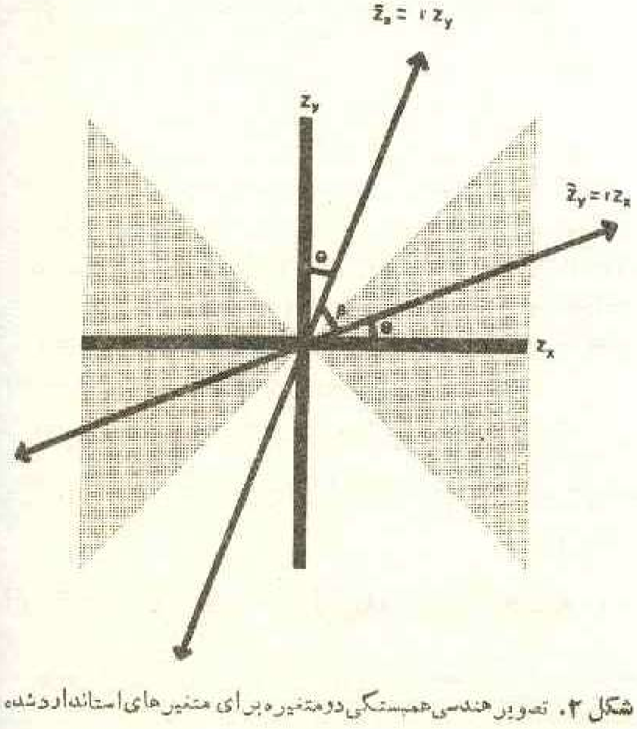
مدل هندسی استاندارد برای نمایش تصویری بستگی بین متغیرها، نمودار پراکنش است. دراین‌مدل، مشاهدات‌را به‌صورت نقطه‌هایی ازفضایی نمایش می‌دهیم که با محورهای متغیرها تعریف می‌شود. صورت "وارونهٔ" این فضا را کــه معمولاً بــه "فضای شخصی" موسوم است، می‌توان با فرض‌اینکه هرمحور مشاهده‌ای را نمایش دهد، تعریف کرد. این‌فضا دونقطه‌را در بردارد ــ یکی برای‌هرمتغیر ــ که نقاط انتهایی بردارهای واقع دراین‌فضای (بالقوه) کلان‌بعدی را تعریف می‌کند. گرچه چندبعدی بودن این فضا مانع تجسم آن است، اما دوبردار متغیر، یك زیرفضای دوبعدی را تعریف‌می‌کنند که به‌آسانی قابل تجسم است.

اگر بردارهای متغیر مبتنی برمتغیرهای مرکزی‌شده باشند، آنگاه همبستگی با α که زاویهٔ بین بردارهای‌متغیر است بستگی سرراست دارد [۲۱]:

$$r=\cos(\alpha) \qquad (۱\cdot۸)$$

وقتی زاویه ۰ است، بردارهــا روی یــك خط می‌افتند و درنتیجه $\cos(\alpha)=\pm۱$. وقتی‌زاویه ۹۰° است، بردارها برهم‌عمودند و $\cos(\alpha)=۰$. (راجرز، نایسواندر، توتاکر [۳۲] بستگی بین بردارهــای متغیر متعامد و ناهمبسته را در فضــای شخصی نشان داده‌اند.)

برای‌تجسم همبستگی، مشاهدهٔ یك‌زاویه خیلی‌ساده‌تر ازمشاهدهٔ چگونگی گردآمدن نقاط حول خط رگرسیون است. به عقیدهٔ ما، این تعبیر درمقایسه با سایرتعبیرها آسانترین راه "ملاحظهٔ" اندازهٔ



شکل ۲. تصویرهندسی همبستگی دومتغیره برای متغیرهای استاندارد شده

همبستگی است؛ زیرا می‌توان مستقیماً اندازهٔ زاویهٔ بین دو بردار را مشاهده کرد. اما، این فضای "وارونه"، کـه اجازه می‌دهد r را به‌صورت کسینوس یک زاویه به‌نمایش دهیم به‌عنوان یک ابزار تعبیر کننده، نسبتاً از نظر دور مانده است. چند تعبیر مربوط به تجزیهٔ عاملی، نمایشهای هندسی تحلیل رگرسیون چندگانهٔ درپیر[1] و اسمیت ([۷] صص ۲۰۱-۲۰۳)، و هک[2] و ساندلر [۱۱، ص۵۲]، از جمله استثناهایی هستند که باید برشمرد. فیشر نیز برای تفهیم بینشهای آمـاری استادانه‌اش کراراً از این فضا استفاده کرده است ([۹] را ببینید).

۹. همبستگی به‌صورت یک واریانس تجدید مقیاس شده از تفاضل بین اندازه‌های استاندارد شده

$z_Y - z_X$ را به عنوان تفاضل بین متغیرهـای استاندارد شدهٔ X و Y برای هر مشاهده تعریف می‌کنیم. در این صورت

$$r = 1 - \frac{S^2_{(z_Y - z_X)}}{2} \qquad (1.9)$$

برای نشان‌دادن این رابطه می‌توانیم با واریانس تفاضلی

$$S^2_{Y-X} = S^2_X + S^2_Y - 2rS_XS_Y$$

شروع کنیم. چون وقتی متغیرها را استانـدارد می‌کنیم، انحراف معیارهـا و واریانسها برابر واحد می‌شوند، می‌تـوانیم به‌آسانی رابطهٔ بـالا را نسبت به r حل کنیم و معادلهٔ (۱.۹) را بـه‌دست آوریم.

توجه به‌این نکته جالب است که در این معادله، چون همبستگی به بازهٔ از ۱- تا ۱ محدود است، واریانس این اندازهٔ تفاضلی به بازهٔ از ۰ تا ۴ محدود می‌شود. بنا بر این واریانس تفاضلی اندازه‌های استاندارد شده هرگز از ۴ تجاوز نمی‌کند. وقتی همبستگی برابر ۱- است، واریانس به‌کران بالایی می‌رسد.

می‌تـوانیم r را به‌صورت واریانس مجموعی از متغیرهـای استاندارد شده نیز تعریف کنیم:

$$r = \frac{S^2_{(z_Y + z_X)}}{2} - 1 \qquad (2.9)$$

در اینجا، واریانس‌مجموع نیز از ۰ تا ۴ تغییر می‌کند، و موقعی که همبستگی برابر ۱+ است به‌کران بـالای خود می‌رسد. مقدار این نهمین تعبیر است که نشان می‌دهد همبستگی، تبدیل خطی نوع خاصی از واریانس است. لذا، با معلوم بودن همبستگی، می‌توانیم مستقیماً واریانس مجموع یا واریانس تفاضل متغیرهای استاندارد شده را تعریف کنیم، و برعکس.

تمام نه‌تعبیر قبلی ضریب همبستگی ماهیتاً جبری و مثلثاتی هستند. تا اینجا دربارهٔ ماهیت توزیعهای تک‌متغیره یا دومتغیره X و Y هیچ فرضی نشد. در تعبیرهـای پایانی، نرمال بودن دومتغیره را فرض خواهیم کرد. توجه‌خود را همچنان به‌صورتهای مفهومی ومحاسباتی r معطوف می‌داریم، امـا چند تعبیر آخری را براین فرض مشترک دربارهٔ توزیع جامعه، استوار می‌کنیم.

1. Draper 2. Huck

۱۰. همبستگی برآورد شده از روی قاعدهٔ بادکنکی

این تعبیرمنسوب به‌شاتیون[1] [۴] است. او پیشنهاد کرد کـه پیرامون نمودار پراکنش یک بستگی دومتغیره، "بادکنکی" رسم کنیم. این بادکنک در واقع یک بیضی تقریبی است که از روی آن دو اندازهٔ h و H به‌دست‌می‌آیند (شکل ۳ را ببینید). h، قطر قائم بیضی‌است که از مرکز توزیع واقع برمحور X می‌گذرد؛ H، برد تغییرات عرضی بیضی روی محور Y است. شاتیون نشان داد که همبستگی را می‌توان به‌طور تقریبی به‌صورت

$$r = \sqrt{1-\left(\frac{h}{H}\right)^2} \qquad (1.10)$$

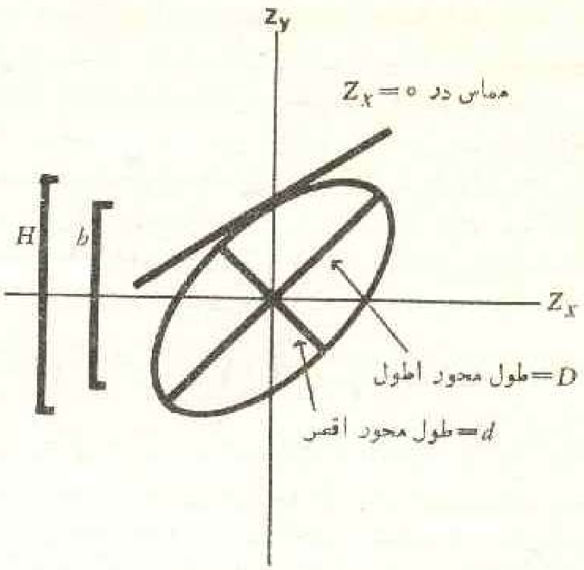
محاسبه کرد. وی با فرض توزیع نرمال دومتغیره و توزیع یکنواخت دومتغیره، دربارهٔ کارایی این‌شیوهٔ محاسباتی تقریبی وسهل‌الوصول، توجیهی نظری ارائه داد. او همچنین مثالهایی چند معرفی کرد که در آنها این‌تکنیک به‌خوبی قابل‌استفاده است. پیشنهاد اغواکننده‌ای که وی عرضه کرد این است که "قاعدهٔ بادکنکی" را می‌توان برای بنای تقریبی یک بستگی دومتغیره با همبستگی‌معینی به‌کاربرد. بدین طریق‌که بیضی‌ای می‌کشیم که r مطلوب را تولید کند و آنگاه سراسر بیضی را به‌طـور یکنواخت از نقطه‌ها پرمی‌کنیم. تـوماس [۲۵] یک "نوموگراف‌جیبی" ساخت که‌نواری "۵ × "۳ است که‌می‌توان از آن برای "مشاهدهٔ" یک بستگی دومتغیره استفاده‌کرد و همبستگی مبتنی برقاعدهٔ بادکنکی را برآورد نمود.

۱۱. همبستگی در ارتباط با بیضیهای دومتغیرهٔ تک‌چگالی

دو نویسندهٔ مختلف، تعبیرهایی از r را که در ارتباط با بیضیهای دومتغیرهٔ تک‌چگالی‌اند پیشنهاد کرده‌اند. توجه‌کنید که این بیضیها صورتهای رسمیتر "بادکنک" بخش ۱۰اند و ساختارهای هندسی هستند که گالتن در داده‌های تجربیش مشاهده کرده است (شکل ۱ را ببینید). شاتیون [۵] رده‌ای از تـوزیعهای دو متغیره (شامل نرمال، یکنواخت، وآمیزه‌هایی از یکنواخت) را ارائه داده است که دارای‌خمهای تک‌چگالی بیضی‌شکل‌هستند. بامعلوم بودن‌همبستگی جامعه‌ای، برای هر ثابت مثبت یک بیضی وجود دارد. بادکنکی که پیرامون یک نمودار پراکنش کشیده می‌شود به ازای ثـابت مثبت بزرگی یکی‌از این بیضیها را تقریب می‌کند. اگرمتغیرها استاندارد شده بـاشند، آنگاه مـرکز این بیضیها در مبدأ است. قطر اطول برای $\rho > 0$ برنیمساز ربع‌اول وسوم و برای $\rho < 0$ برنیمساز دوربع دیگر می‌افتد.

مارکس [۱۴]، با محاسبه‌ای‌ساده نشان دادکه‌شیب خط مماس در $z_X = 0$، برابر با همبستگی است. شکل ۳ این خط مماس را نشان‌می‌دهدکه شیب آن مساوی با r است. وقتی‌همبستگی ۰ است، بیضی یک‌دایره است و مماس دارای شیب ۰ است. وقتی‌همبستگی ۱ است، بیضی به‌خطی‌مستقیم میل‌می‌کند که همان نیمساز (با شیب ۱) است. توجه‌کنید که چون تمام بیضیهای تک‌چگالی موازی‌اند، تعبیر ارائه شده به‌انتخاب بیضی‌بستگی ندارد. همچنین شایان توجه است که شیب خط مماس در $z_X = 0$ همان شیب خط رگرسیون

1. Chatillon



شکل ۳. همبستگی در ارتباط با تابعهایی از بیضیهای تک چگالی

استاندارد شده است (بخش ۳ را ببینید).

شیلینگ [۲۳] نیز برای رسیدن به بستگی مشابهی، از این چارچوب استفاده کرده است. فرض کنید متغیرها استاندارد شده باشند به‌قسمی که، مثل قبل، مرکز بیضیها در مبدأ باشد. اگر D، درازای قطر اطول یکی از بیضیهای تک‌چگالی و d، درازای قطر اقصر آن باشد، آنگاه

$$r=\frac{(D^2-d^2)}{(D^2+d^2)} \qquad (1.11)$$

این محورها نیز در شکل ۳ رسم شده‌اند، و تعبیر، مثل قبل، به‌انتخاب بیضی بستگی ندارد.

۱۲. همبستگی به‌صورت تابعی از آمارهٔ آزمونی از آزمایشهای طرح‌شده

تعبیرهای قبلی r بر متغیرهای کمی مبتنی بودند. دوازدهمین نمایش همبستگی، بستگی آن را با آمارهٔ آزمونی از آزمایشهای طرح‌شده نشان می‌دهد که در آنها یکی از متغیرها (متغیر مستقل)، متغیری رسته‌ای است. این تعبیر، ساختگی بودن تمایز آزمایشها از همبستگی را در بحث طرح آزمایشها ثابت می‌کند. درواقع، فیشر (۱۹۵۲) در اصل، تحلیل واریانس را برحسب ضریب همبستگی درون‌رده‌ای ارائه داد (ر.ك. [۱]).

فرض می کنیم آزمایش طرح شده‌ای با دوشرط تیماری داشته باشیم. مدل آماری استاندارد برای آزمودن تفاوت بین شرایط، آزمون tی مربوط به دو نمونهٔ مستقل است. اگر X به صورت یك متغیر دوحالتی معرف عضویت گروه، تعریف شود (۰، اگر گروه ۱، ۱، اگر گروه ۲)، آنگاه همبستگی بین X و متغیر وابستهٔ Y عبارت است از

$$r=\frac{t}{\sqrt{t^2+n-2}} \qquad (1.12)$$

که در آن، n تعداد کل مشاهدات دو گروه تیماری است. این ضریب همبستگی را می‌توان به‌عنوان معیاری برای قدرت یك اثر تیماری، درمقابل معنادار بودن یك اثر به‌کار برد. در این وضعیت آزمون معنادار بودن r، همان آزمون tی معمولی را به دست می‌دهد. بنابراین، r به‌وضوح می‌تواند به‌همان خوبی که معیاری از پیوند را در وضعیتهای مشاهداتی فراهم می کند به عنوان یك آمارهٔ آزمون نیز در یك آزمایش طرح شده به‌کار رود.

در حالت تحلیل واریانس با گروههای بیشتر یا عاملهای چندگانه، تعمیم این بستگی، ضریبهای همبستگی چندگانهٔ مربوط به‌اثرهای اصلی و اثرهای متقابل در آزمایشهای پیچیده‌تر را تعریف می کند. مثلاً، در حالت تحلیل واریانس یکطرفه با k گروه و مجموع کل N آزمودنی، مربع همبستگی چندگانه بین متغیر وابسته و ستونهای ماتریس طرح، ازطریق فرمول زیر به‌آمارهٔ F مربوط می‌شود [۷، ص ۹۳]:

$$R^2=\frac{F(k-1)}{[F(k-1)+(N-k)]}$$

۱۳. همبستگی به‌صورت نسبت دو میانگین

این سومین تعبیر همبستگی است که در ارتباط با میانگینهاست (بخشهای ۴ و ۶ را ببینید). این تعبیر، پایانی مناسب بر مقالهٔ ماست، زیرا اولین بار گالتن آن را پیشنهاد کرده است. به‌علاوه، قدیمترین استنباط گالتن از همبستگی و محاسبهٔ آن، بر این تعبیر مبتنی بود. نایسواندر و پرایس [۱۷] صورت پیشرفته‌تری از این تعبیر را ارائه داده‌اند.

اما گالتن، طبیعی بود که توجه خود را بر همبستگی به‌صورت نسبت میانگینها متمرکز کند، زیرا وی به سؤالهایی از این قبیل علاقه‌مند بود که چگونه باید متوسط قد پدرهایی را که به‌طور غیرمعمول بلندقدند با متوسط قد پسرانشان مقایسه کرد. در بحث زیر به جای نماد نمونه‌ای، از نماد جامعه‌ای استفاده می‌شود، زیرا این تنها درحد (حجم نمونه‌ای افزایشی) است که عبارت نسبت میانگینها همان مقادیر r پیرسن را به‌دست می‌دهد.

وضعیتی مشابه آنچه مورد توجه گالتن بود درنظر می گیریم. فرض کنید X متغیری باشد که بهرهٔ هوشی مادر را نشان دهد، و Y متغیری باشد که بهرهٔ هوشی بزرگترین فرزند این مادر را نمایش دهد. به‌علاوه فرض کنید که $\mu(X)$ و $\mu(Y)$ برابر ۰ و انحراف معیارهای $\sigma(X)$ و $\sigma(Y)$ برابر یك باشند. حال مقداری به‌دلخواه بزرگ از X (مثل X_c) را بر می گزینیم، و میانگین بهرهٔ هوشی مادرهایی را که بهرهٔ هوشی آنها از X بزرگتر است محاسبه می کنیم. این میانگین را با $\mu(X|X>X_c)$ نشان می‌دهیم ــ این، متوسط بهرهٔ هوشی مادرانی است که بهرهٔ هوشی آنها بزرگتر از X_c است. سپس متوسط اندازه‌های Y، یعنی بهرهٔ هوشی بزرگترین فرزند این مادرهای خاص را محاسبه می کنیم. این میانگین را با $\mu(Y|X>X_c)$ نمایش می‌دهیم ــ یعنی، متوسط بهرهٔ هوشی بزرگترین فرزند مادرانی که بهرهٔ هوشی آنها بزرگتر از X_c است. در این صورت، می‌توان نشان داد که

$$r=\frac{\mu(Y|X>X_c)-\mu_Y}{\mu(X|X>X_c)-\mu_X}=\frac{\mu(Y|X>X_c)}{\mu(X|X>X_c)}. \qquad (1.13)$$

برای اثبات (۱.۱۳)، لازم است که توزیع توأم دومتغیر استاندارد شدهٔ X و Y را نرمال فرض کنیم. برهـان سرراست است و تنها مبتنی براین واقعیت است که برای z_X و z_Y، مقدار r برابر بـا شیب‌خط رگرسیون و همچنین برابربا نسبت این دومیانگین‌شرطی است.

مثال ما، مثال ویژه‌ای است، اما این تعبیر دربارهٔ هروضعیتی که روی یك‌متغیر انتخاب صریح رخ می‌دهد؛ ومتغیردوم هم به‌طور ضمنی‌انتخاب‌می‌شود کاربرد دارد. بروگدن[۲] تعبیر نسبت‌میانگینها را به‌کاربرد تانشان‌دهد که وقتی یك آزمون روانشناسی برای‌انتخاب کارکنان مورد استفاده قرار می گیرد، همبستگی بین اندازهٔ آزمون و اندازهٔ ملاك، معیار مناسبی را به‌دست می‌دهد کـه بیان می کند آزمون تـا چه حد ابزار انتخاب "کاملسی" است. محققاً امـکان استفاده‌های دیگری از این تعبیر وجود دارد.

سخن‌آخر

محققاً راههای دیگری‌هم برای تعبیرضریب همبستگی وجود دارد. وقتی برای‌مسألهٔ همبستگی رهیافتی اختیار می‌شودکه بیشترآمادی وکمترجبری است،روشهای عدیدهٔ نموداری مفید و جالب دیگری موجودند. ما به‌هیچ‌وجه براین عقیده نیستیم که حتی درچارچوب نسبتاً محدودی کـه برای کارمان اختیار کرده‌ایم تمام رهیافتهای مفید یـا جالب را خلاصه نموده‌ایم، مع‌هذا، این ۱۳ رهیافت، تنوع تعبیرهای موجود را به‌معلمین ومحققینی که همبستگی را به‌کار می‌برند نشان می‌دهد.

کار اولیهٔ گالتن در زمینهٔ همبستگی، از یك مسألهٔ زیست‌سنجی بسیارخاص‌نشأت گرفت. جالب‌است که‌چنین کوششی که‌برموضوعی خاص متمرکز بود، به‌چیزی منجرشد که شاید متداولترین شاخص عملی در رشتهٔ آمار باشد. دامنهٔ تعبیرهای ضریب همبستگی، رشد این شاخص مهم را درطول یك قرن نشان می‌دهد. ازسوی دیگر، شاخص گالتن و پیرسن، به‌گونه‌ای شگفت‌انگیز، به همان صورت که دراصل پیشنهاد شده است بدون تغییرباقی مانده است.

مراجع

1. Box, J. F. (1978), *R. A. Fisher: The Life of a Scientist*, New York: John Wiley.
2. Brogden, H. E. (1946), "On the Interpretation of the Correlation Coefficient as a Measure of Predictive Efficiency," *Journal of Educational Psychology* , 37, 65-76.
3. Carroll, J.B. (1961), "The Nature of the Data, or How to Choose a Correlation Coefficient," *Psychometrika*, 26, 347-372.
4. Chatillon, G. (1984 a), "The Balloon Rules for a Rough Estimate of the Correlation Coefficient," *The American Statistician*, 38, 58-60.
5. ——(1984 b), "Reply to Schilling," *The American Statistician*, 38,330.
6. Cook, T.C., and Campbell, D.T. (1979), *Quasi - Experimentation*, Boston: Houghton Mifflin.
7. Draper, N. R. and Smith, H. (1981), *Applied Regression Analysis*, New York: John Wiley.
8. Fisher, R. A. (1925), *Statistical Methods for Research Workers*, Edinburgh, U.K.: Oliver & Boyd.
9. Galton, F. (1885), "Regression Towards Mediocrity in Hereditary Stature," *Journal of the Anthropological Institute*, 15, 246-263.
10. Henrysson, S. (1971), "Gathering, Analyzing, and Using Data on Test Items," in *Educational Measurement*, ed. R.L. Thorndike, Washington, DC: American Council on Education, pp. 130-159.
11. Huck, S.W., and Sandler, H.M. (1984), *Statistical Illusions: Solutions*, New York: Harper & Row.
12. Lord, F.M., and Novick, M.R. (1968), *Statistical Theories of Mental Test Scores*, Reading, MA: Addison - Wesley.
13. MacKenzie, D.A. (1981), *Statistics in Britain: 1865-1930*, Edinburgh, U.K.: Edinburgh University Press.
14. Marks, E. (1982), "A Note on the Geometric Interpretation of the Correlation Coefficient," *Journal of Educational Statistics*, 7, 233-237.
15. Mulaik, S.A. (1972), *The Foundations of Factor Analysis*, New York: McGraw-Hill.
16. Nefzger, M.D., and Drasgow, J. (1957), "The Needless Assumption of Normality in Pearson's *r*," *The American Psychologist*, 12, 623-625.
17. Nicewander, W.A., and Price, J.M. (1982), "The Correlation Coefficient as the Ratio of Two Means: An Interpretation Due to Galton and Brogden," unpublished paper presented to the Psychometric Society, Montreal, Canada, May.
18. Ozer, D.J. (1985), "Correlation and the Coefficient of Determination, " *Psychological Bulletin*, 97, 307-315.
19. Pearson, K. (1895), *Royal Society Proceedings*, 58, 241.
20. —— (1920), "Notes on the History of Correlation," *Biometrika*, 13, 25-45.
21. Rodgers, J.L. (1982), "On Huge Dimensional Spaces: The Hidden Superspace in Regression and Correlation Analysis," unpublished paper presented to the Psychometric Society, Montreal, Canada, May.
22. Rodgers, J.L., Nicewander, W.A. and Toothaker, L. (1984), "Linearly Independent, Orthogonal, and Uncorrelated Variables," *The American Statistician*, 38, 133-134.
23. Schilling, M.F. (1984), "Some Remarks on Quick Estimation of the Correlation Coefficient," *The American Statistician*, 38, 330.
24. Seal, H.L. (1967), "The Historical Development of the Gauss Linear Model," *Biometrika*, 54, 1-24.
25. Thomas, H. (1984), "Psychophysical Estimation of the Correlation Coefficient From a Bivariate Scatterplot: A Pocket Visual Nomograph," unpublished paper presented to the Psychometric Society, Santa Barbara, California, June.
26. Walker, H.M. (1929), *Studies in the History of Statistical Method*, Baltimore: Williams & Wilkins.

• Joseph Lee Rodgers, W. Alan Nicewander, "Thirteen ways to look at the correlation coefficient," *The American Statistician*, (1) **42** (1988) 59-66.

★ جوزف راجرز و الن نایسواندر، بخش روانشناسی دانشگاه اوکلاهمای امریکا