

بازگشت به میانگین در مطالعات اپیدمیولوژیک

سیما مسعودی^۱، یدالله محرابی^۲، داود خلیلی^۳، پروین یآوری^۴

^۱ دکتری اپیدمیولوژی، دانشگاه علوم پزشکی ارومیه و دانشکده بهداشت دانشگاه علوم پزشکی شهید بهشتی

^۲ استاد آمار زیستی دانشکده بهداشت دانشگاه علوم پزشکی شهید بهشتی

^۳ استادیار اپیدمیولوژی پژوهشکده غدد درون‌ریز و متابولیسم دانشگاه علوم پزشکی شهید بهشتی

^۴ استاد اپیدمیولوژی، گروه بهداشت و پزشکی اجتماعی دانشگاه علوم پزشکی شهید بهشتی، مرکز تحقیقات ژنومیک و انجمن اپیدمیولوژیست‌های ایران

نویسنده رابط: پروین یآوری، نشانی: تهران، ولنجک، بلوار دانشجو، دانشکده پزشکی دانشگاه علوم پزشکی شهید بهشتی، ساختمان شماره ۱، طبقه هشتم، گروه پزشکی اجتماعی.

تلفن: ۲۲۴۳۹۹۳۶-۰۲۱، پست الکترونیک: p.yavari-grc@sbmu.ac.ir

تاریخ دریافت: ۹۳/۰۵/۰۱؛ پذیرش: ۹۳/۰۸/۰۳

مقدمه و اهداف: اندازه‌گیری ویژگی‌های مورد بررسی در تحقیقات علوم پزشکی تقریباً همیشه در معرض خطای اندازه‌گیری تصادفی قرار دارند که معمولاً این خطا و اثرات ناشی از آن توسط پژوهشگران نادیده گرفته می‌شود. یکی از اثرهای خطای تصادفی در اندازه‌گیری یک پدیده آماری به نام «بازگشت به میانگین» است. این پدیده زمانی مشاهده می‌گردد که بر اساس اندازه یک متغیر، گروهی انتهایی از جمعیت انتخاب گردد. اگر در این گروه اندازه‌گیری دومی از متغیر انجام شود. در مقایسه با میانگین اندازه‌گیری اول، میانگین اندازه‌گیری دوم به میانگین جمعیت نزدیک‌تر خواهد شد. در مطالعه‌های مداخله‌ای، این افزایش (یا کاهش) ممکن است اثر مداخله انجام شده تلقی شده و نتیجه گرفته شود که درمان مؤثر بوده است، در حالی که اثری نداشته است. نادیده گرفتن بازگشت به میانگین سبب خطا در نتیجه‌گیری و تفسیر نتایج مطالعه‌ها، تصمیم‌گیری در پزشکی مبتنی بر شواهد و برنامه‌ریزی برای اقدامات پیشگیری و بهداشتی خواهد شد. این مقاله به اهمیت این موضوع و اثر آن در مطالعه‌های مداخله‌ای اپیدمیولوژیک و راه‌های اجتناب از آن می‌پردازد.

واژگان کلیدی: خطای اندازه‌گیری تصادفی، بازگشت به میانگین، مطالعه اپیدمیولوژیک

مقدمه

اندازه‌گیری شونده است که هر یک از این مؤلفه‌ها می‌توانند منبعی برای خطای اندازه‌گیری باشند. به طور کلی با توجه به ماهیت خطاها، می‌توان آن‌ها را در دو گروه قرار داد: خطاهای سیستماتیک و خطاهای تصادفی. خطاهای سیستماتیک یا منظم خطاهایی هستند که مقدار آن‌ها در هر بار اندازه‌گیری یا ثابت می‌مانند یا این که همیشه در یک جهت معین تغییر می‌کنند، یعنی همیشه بیش‌تر (یا کم‌تر) از مقدار واقعی هستند؛ بنابراین برای خطا قابل پیش‌بینی است و در اندازه‌گیری‌های مکرر از یک ویژگی، مقدار میانگین خطاها، صفر نخواهد بود (۵،۴). در خطای تصادفی مقدار و جهت خطا در هر بار اندازه‌گیری، غیرقابل پیش‌بینی و تصادفی است، اما در تکرار اندازه‌گیری به دفعات خیلی زیاد مقدار میانگین خطاها به صفر رسیده و میانگین مقادیر مشاهده شده با مقدار واقعی متغیر برابر خواهد بود، به عبارتی خطاها همدیگر را خنثی می‌کنند (۴). این موضوع سبب نادیده گرفتن خطای تصادفی توسط پژوهشگران در بیش‌تر پژوهش‌ها می‌شود. با این حال در مطالعه‌های بالینی که زیرگروه خاصی از جمعیت مورد مطالعه قرار می‌گیرد یا در مطالعه‌های علیتی که اثر

در تمامی مطالعه‌هایی که در هر رشته‌ای از علوم از جمله در پزشکی و اپیدمیولوژی انجام می‌گیرد، اندازه‌گیری ویژگی‌های افراد (متغیرهای مطالعه) از اهمیت خاصی برخوردار بوده و اعتبار یافته‌های پژوهش به میزان زیادی به صحت و دقت اندازه‌گیری این متغیرها بستگی دارد، اما همه اندازه‌گیری‌ها در معرض خطا قرار دارند. خطا در اندازه‌گیری متغیرهای کیفی را سوءطبقه‌بندی^۱ و خطا در اندازه‌گیری متغیرهای کمی را خطای اندازه‌گیری^۲ می‌نامند (۱). در سراسر این مقاله، خطای اندازه‌گیری در متغیرهای کمی مورد نظر است. در تکرار اندازه‌گیری یک متغیر در یک فرد وجود خطا در اندازه‌گیری سبب می‌شود که یافته‌های مشاهده شده در اندازه‌گیری‌ها با هم تفاوت داشته باشند؛ بنابراین خطای اندازه‌گیری به صورت تفاوت بین مقدار مشاهده شده (اندازه‌گیری شده) و مقدار واقعی متغیر تعریف می‌شود (۲،۳). در مطالعه‌های اپیدمیولوژیک هر اندازه‌گیری دارای ۴ مؤلفه ابزار اندازه‌گیری، روش اندازه‌گیری، فرد اندازه‌گیری کننده و فرد

^۱ Misclassification

^۲ Measurement error

اندازه‌گیری معین چه مقداری مشاهده خواهد شد. اگر مقدار خطای اندازه‌گیری زیاد باشد؛ واریانس خطا بزرگ شده و شکل توزیع پهن خواهد شد و برعکس اگر خطا کم باشد؛ واریانس خطای اندازه‌گیری کوچک بوده و شکل توزیع باریک‌تر خواهد شد. در حالت ایده‌آل اگر هیچ خطایی در اندازه‌گیری رخ ندهد، مقادیر مشاهده شده با مقدار واقعی متغیر برابر خواهند بود (۶).

فرض کنید جمعیتی از افراد وجود دارد که توزیع متغیر مورد نظر در این جمعیت نرمال است. این توزیع از مقادیر همه افراد جمعیت تشکیل می‌شود، و مقدار واقعی متغیر برای هر فرد (T_i) ثابت است و مقدار مشاهده شده برای هر فرد (X_i) تنها یکی از مقادیر توزیع فردی متغیر است، یعنی:

$$X_i = T_i + e_i \quad i = 1, \dots, n, \quad e_i \sim N(0, \sigma_{e_i}^2)$$

میانگین توزیع جمعیتی متغیر برای مقادیر مشاهده شده برابر است با.

$$\mu_X = \frac{\sum X_i}{n} = \frac{\sum (T_i + e_i)}{n} = \frac{\sum T_i + \sum e_i}{n} = \frac{\sum T_i}{n} = \mu_T$$

یعنی خطاهای اندازه‌گیری تصادفی در سطح جمعیت اثر یکدیگر را خنثی می‌کنند و در نتیجه میانگین مقادیر مشاهده شده برای جمعیت برابر میانگین واقعی متغیر است، اما واریانس مقادیر مشاهده شده بزرگ‌تر می‌شود زیرا شامل واریانس مقادیر واقعی و واریانس خطای اندازه‌گیری است ($\sigma_X^2 = \sigma_T^2 + \sigma_e^2$). به عبارتی وجود خطای تصادفی در اندازه‌گیری، تغییرپذیری مقادیر متغیر را افزایش می‌دهد.

در علوم بالینی خطای اندازه‌گیری تصادفی می‌تواند ناشی از ابزار اندازه‌گیری، فرد اندازه‌گیری کننده و یا نوسانات واقعی کوتاه مدت در متغیر مورد بررسی باشد. کالیبره کردن ابزار، استفاده از ابزارهای دقیق و پروتکل‌های استاندارد و آموزش افرادی که اندازه‌گیری را انجام می‌دهند، می‌تواند خطاهای ناشی از ابزار و افراد را به حداقل برساند، اما آن را کاملاً حذف نمی‌کند. گذشته از این بسیاری از متغیرهای مورد بررسی در علوم پزشکی و بهداشتی، به دلیل نوسانات بیولوژیک همیشه مقدار ثابتی نداشته و اندازه آن‌ها از زمانی به زمان دیگر تغییر می‌کند. در عمل مقدار واقعی هر متغیر بیولوژیک به عنوان میانگین اندازه‌گیری‌های مکرر در طول زمان تعریف می‌گردد (۲). برای مثال فشار خون تحت تأثیر عوامل مختلفی از جمله تغذیه، فعالیت فیزیکی، سن، وضع روحی فرد و بسیاری عوامل دیگر قرار دارد. اگر فشار خون شخصی به دفعات در طی چندین ماه یا حتی چند سال اندازه‌گیری شود، مقادیر یکسانی در همه اندازه‌گیری‌ها به دست نخواهد آمد، اما میانگین

عوامل خطر روی رخداد بیماری‌ها بررسی می‌شود، وجود خطای اندازه‌گیری تصادفی در اندازه‌گیری متغیرها و نادیده گرفتن آن در طراحی و تجزیه و تحلیل یافته‌های مطالعه‌ها ضمن لطمه زدن به اعتبار مطالعه می‌تواند به نتیجه‌گیری‌ها، تفسیرها و تصمیم‌گیری‌های نادرست منجر گردد.

خطای اندازه‌گیری تصادفی

همان‌طور که گفته شد خطاهای اندازه‌گیری تصادفی غیرقابل پیش‌بینی بوده و می‌توانند در هر جهتی رخ بدهند، اما اگر اندازه‌گیری به دفعات زیادی تکرار گردد، این خطاها همدیگر را خنثی کرده و میانگین مقادیر مشاهده شده متغیر با میانگین واقعی آن برابر خواهد بود. در این جا ابتدا این مفهوم بیش‌تر مورد بررسی قرار می‌گیرد. در سراسر این مقاله فرض بر این است که در اندازه‌گیری‌ها، خطای سیستماتیک وجود ندارد.

هر ویژگی یا متغیر در هر فردی دارای یک مقدار واقعی است که در اندازه‌گیری‌های متعدد این مقدار ثابت است، اما وجود خطای تصادفی در اندازه‌گیری سبب می‌گردد که مقدار مشاهده شده متغیر با مقدار واقعی آن متفاوت باشد. به عبارتی می‌توان نوشت:

$$X = T + e$$

در این مدل X مقدار مشاهده شده، T مقدار واقعی و e مقدار خطای اندازه‌گیری است و فرض می‌شود که خطای اندازه‌گیری دارای توزیع نرمال با میانگین صفر و واریانس σ_e^2 است. این خطا سبب می‌شود که مقادیر مشاهده شده در اندازه‌گیری‌های تکراری از یک متغیر با هم برابر نباشند. در سطح فردی تغییر در مقادیر مشاهده شده، «تغییرپذیری درون فردی»^۱ نامیده می‌شود. اگر متغیری به دفعات زیادی اندازه‌گیری شود، می‌توان نمودار توزیع آن را برای الگوی تغییرات مشاهده شده ترسیم کنیم. به این ترتیب مقدار مشاهده شده در هر اندازه‌گیری یک نمونه تصادفی از توزیع این متغیر است. در تئوری کلاسیک آزمون فرض می‌شود که توزیع مقادیر مشاهده شده برای متغیر نرمال است و میانگین این توزیع برابر مقدار واقعی متغیر برای فرد و واریانس آن واریانس خطای اندازه‌گیری است. اگر این توزیع معلوم باشد، به راحتی می‌توان احتمال مشاهده مقداری برابر یا بیش‌تر (یا کم‌تر) از یک مقدار معین از متغیر را به دست آورد، اما حتی اگر مقدار واقعی متغیر برای فردی معلوم باشد؛ نمی‌توان تعیین نمود که در یک

^۱ Within person variability

اندازه‌گیری‌های مکرر به فشارخون واقعی فرد نزدیک‌تر خواهد بود.

بازگشت به میانگین چیست و چه زمانی رخ می‌دهد

در بخش قبل نشان داده شد که چگونه خطاهای تصادفی که در سطح فردی رخ می‌دهند؛ در سطح جمعیت یکدیگر را خنثی می‌کنند و در نتیجه میانگین مقادیر مشاهده شده برابر میانگین واقعی متغیر در جمعیت می‌شود. حال فرض کنید متغیری برای همه افراد یک جمعیت با یک فاصله زمانی معین (مثلاً چند روز یا چند هفته) دو بار اندازه‌گیری شود. اگر هیچ خطایی در اندازه‌گیری‌ها رخ ندهد، مقادیر به‌دست آمده برای هر فرد در دو اندازه‌گیری یکسان و ضریب همبستگی بین دو اندازه برابر یک خواهد شد (یک همبستگی عالی). در عمل به دلیل خطای تصادفی، مقادیر مشاهده شده در دو اندازه‌گیری با هم یکسان نخواهد بود و در نتیجه ضریب همبستگی بین دو اندازه از یک کم‌تر خواهد شد. در چنین شرایطی افرادی که در اندازه‌گیری اول مقادیر بالایی از توزیع جمعیتی مقادیر متغیر را داشتند، مقدار متغیر در اندازه‌گیری دوم کم‌تر بوده و به میانگین توزیع نزدیک خواهد شد. افرادی هم که در اندازه‌گیری اول مقادیر پایینی داشته‌اند، مقدار متغیر برای آن‌ها در اندازه‌گیری دوم بیش‌تر از مقدار اول بوده و به میانگین توزیع نزدیک خواهد شد. به طور کلی در اندازه‌گیری اول مقادیر بزرگتر عموماً دارای خطاهای اندازه‌گیری مثبت و مقادیر کوچکتر دارای خطای اندازه‌گیری منفی هستند. این وضع در دو انتهای توزیع که دارای مقادیر خیلی بزرگ و کوچک هستند، بیش‌تر دیده می‌شود. در کل اندازه‌گیری دوم به میانگین توزیع نزدیک می‌شود، و سبب می‌گردد که بین مقادیر اندازه‌گیری اول و تفاوت بین دو اندازه یک رابطه منفی وجود داشته باشد (۷). تغییرات فردی ذکر شده در سطح جمعیت یکدیگر را خنثی کرده و میانگین دو اندازه‌گیری با هم برابر یا نزدیک به هم بوده و تفاوت بین دو میانگین تنها به علت شانس خواهد بود.

در بیش‌تر مطالعه‌های اپیدمیولوژی در حیطه پزشکی و بهداشت، به جای کل جمعیت، یک زیرگروه انتخابی از آن مورد مطالعه قرار می‌گیرد، و معیار واجد شرایط بودن برای شرکت در مطالعه، داشتن سطحی خاص از یک متغیر کمی است، و افراد بیمار و غیر بیمار توسط یک حد آستانه قراردادی از هم جدا می‌شوند. بیمارانی که به این ترتیب برای شرکت در مطالعه

انتخاب می‌شوند؛ دارای یک توزیع قیچی شده^۱ از جمعیت هستند. برای مثال کلسترول بالای ۲۶۰ میلی‌گرم در دسی‌لیتر و HDL کم‌تر از ۴۰ میلی‌گرم در دسی‌لیتر به عنوان عامل خطر بیماری‌های قلبی-عروقی تعیین شده است (۸) یا افرادی مبتلا به پرفشاری خون تشخیص داده می‌شوند که فشارخون سیستولی بیش از ۱۴۰ میلی‌متر جیوه و یا فشارخون دیاستولی بیش از ۹۰ میلی‌متر جیوه داشته باشند (۹).

فرض کنید برای انتخاب زیرگروهی از افراد جمعیت که در آن‌ها مقدار متغیر مورد نظر بالای یک حد آستانه معین باشد، از یک آزمون غربالگری استفاده می‌شود. هم‌چنین فرض کنید که به طریقی مقدار واقعی متغیر برای هر فردی وجود داشته باشد؛ در این صورت در آزمون غربالگری ۴ گروه در اندازه‌گیری وجود خواهند داشت (۱۰) که در جدول شماره ۱ نشان داده شده‌اند.

جدول شماره ۱- قاعده انتخاب افراد برای ورود به مطالعه بر اساس مقادیر واقعی و مشاهده شده آن‌ها

مقدار متغیر در غربالگری	مقدار واقعی متغیر	
	زیر آستانه	بالای آستانه
زیر آستانه	الف وارد مطالعه نمی‌شوند	ب وارد مطالعه نمی‌شوند
بالای آستانه	ج وارد مطالعه می‌شوند	د وارد مطالعه می‌شوند

منبع: استرینر (۱۰)

از این چهار گروه، افراد گروه‌های ج و د که مقدار متغیر برای آن‌ها در غربالگری بالای حد آستانه بوده، وارد مطالعه می‌شوند و افراد گروه الف و ب که در غربالگری مقدار متغیر در آن‌ها کم‌تر از آستانه بوده، وارد مطالعه نخواهند شد. دو گروه ج و د شامل افرادی هستند که حد آستانه انتخابی در دامنه توزیع فردی مقادیر متغیر برای آن‌ها قرار داشته است. مقدار واقعی متغیر برای افراد گروه ج زیر آستانه بوده، اما به دلیل خطای تصادفی در اندازه‌گیری، در غربالگری، مقداری بیش‌تر از حد آستانه را داشته و وارد مطالعه شده‌اند. عکس این حالت برای افراد گروه ب رخ داده است. مقدار واقعی متغیر در این افراد بالای حد آستانه بوده، اما به

^۱ Truncated distribution

تاریخچه

تاریخچه بازگشت به میانگین به قرن نوزدهم و کارهای فرانسیس گالتون روی وراثت قد باز می‌گردد (۱۵). گالتون در ابتدا به مطالعه وراثت هوش علاقه‌مند بود، اما از آن‌جا که ابزار مناسبی برای اندازه‌گیری دقیق هوش افراد وجود نداشت، توجه خود را به قد افراد معطوف داشت. برای این کار وی از خانواده‌هایی که حداقل ۶ نفری بودند برای شرکت در مطالعه دعوت نمود، و به این ترتیب در مطالعه خود قد ۹۳۰ فرد بزرگسال را با قد ۲۰۵ نفر والدین آن‌ها مورد بررسی قرار داد. در محاسبه میانگین قد والدین با توجه به این که زنان به طور معمول کوتاه‌تر از مردان هستند، برای از بین بردن هر گونه تفاوت جنسی در قد والدین، وی قد زنان را در عاملی به اندازه ۱/۰۸ ضرب نمود که معادل با اضافه نمودن کمی کم‌تر از یک دوازدهم قد هر زن بود. سپس والدین را به گروه‌های مساوی قد تقسیم‌بندی و قد فرزندان را در هر گروه به‌دست آورد. وی یک نمودار پراکنش از قد فرزندان را در مقابل قد والدین‌شان ترسیم نمود و متوجه شد اگرچه فرزندان والدین بلندقد به طور معمول بلندقد بودند، اما کوتاه‌تر از والدین خودشان بودند و در مقابل فرزندان والدین کوتاه قد اگرچه کوتاه قد بودند، اما از والدین خوشان بلندتر بودند. در نتیجه قد فرزندان هر دو گروه از والدین به میانگین قد جامعه نزدیک‌تر بود. وی این موضوع را «بازگشت به حد وسط»^۳ نامید که بعدها بازگشت به میانگین نامیده شد. گالتون در این مطالعه رابطه بین اندازه‌های تکراری یا اندازه‌های مربوط به دو متغیر از یک فرد را بررسی نکرد، بلکه آن‌چه که بررسی نمود رابطه بین قد والدین و فرزندان آن‌ها بود، که در آن علاوه از خطاهای اندازه‌گیری، وراثت و عوامل ژنتیکی و همچنین عوامل محیطی در بازگشت به میانگین نقش دارند.

تعیین مقدار بازگشت به میانگین و عوامل مؤثر بر آن

فرض کنید در جمعیتی از افراد میانگین متغیری برابر μ و انحراف معیار آن σ باشد. اگر از این جمعیت نمونه‌ای از افراد انتخاب شوند، که مقدار این متغیر در آن‌ها بیش‌تر از یک حد انتخابی مانند k باشد، در این نمونه انتخابی مقدار میانگین متغیر بیش‌تر از میانگین متغیر در جمعیت خواهد بود. به شرط برقرار بودن فرض‌های زیر می‌توان اثر بازگشت به میانگین را محاسبه نمود:

۱- خطاهای تصادفی هم در افراد مختلف و هم در درون هر فرد

دلیل خطای تصادفی در اندازه‌گیری مقدار مشاهده شده کم‌تر از آستانه بوده و در نتیجه وارد مطالعه نشده‌اند.

اگر پس از مدتی این متغیر را در گروه انتخاب شده برای مطالعه دوباره اندازه‌گیری کرده و نتایج دو اندازه‌گیری با هم مقایسه شود، مشاهده می‌شود که مقدار میانگین متغیر مورد بررسی در اندازه‌گیری دوم به میانگین جمعیتی که گروه از آن انتخاب شده، نزدیک شده است. چرا این امر رخ می‌دهد؟ علت تفاوتی است که بین مقادیر مشاهده شده در دو اندازه‌گیری وجود دارد. افراد گروه د افرادی از انتهای توزیع متغیر هستند که در درصد عمده‌ای از آن‌ها مقدار مشاهده شده در اندازه‌گیری اول بالاتر از میانگین واقعی آن‌ها بود، اما در اندازه‌گیری دوم مقداری کم‌تر از اندازه‌گیری اول را دارند. همچنین سهمی از افراد گروه ج کسانی هستند که مقدار واقعی متغیر زیر حد آستانه انتخابی بوده است. در اندازه‌گیری دوم مقدار مشاهده شده برای آنها کم‌تر از حد آستانه خواهد بود. در نتیجه در اندازه‌گیری دوم میانگین متغیر در این گروه کاهش یافته و به میانگین جمعیت نزدیک می‌شود (۱۰). اگر افراد گروه الف و ب که در غربالگری دارای شرایط شرکت در مطالعه نشدند، وارد مطالعه شده بودند؛ افزایش مقادیر متغیر برای آن‌ها در اندازه‌گیری دوم اثر کاهش در گروه ج و د را خنثی می‌کرد. به طور کلی اگر به دلیل خطاهای تصادفی یا غیرسیستماتیک (ناشی از تکنیک اندازه‌گیری یا تغییرات بیولوژیک در متغیر یا هر دو) در اندازه یک متغیر تغییرپذیری وجود داشته باشد، که سبب شود همبستگی بین اندازه‌گیری‌های تکراری در یک نمونه تصادفی از جمعیت از ۱ کم‌تر شود، در یک زیرگروه انتخابی از این جمعیت (مثلاً بر اساس یک حد آستانه‌ای معین) معمولاً میانگین متغیر در اندازه‌گیری دوم به میانگین جمعیت نزدیک‌تر می‌شود. این پدیده را «بازگشت به میانگین» می‌نامند (۱۴-۱۱). اگرچه بازگشت به میانگین عموماً برای اندازه‌گیری‌های تکراری از یک متغیر در یک گروه انتهایی از جمعیت تعریف می‌شود، اما از نظر بلند^۱ و آلتمن^۲ (۱۱،۱۲) زمانی بازگشت به میانگین رخ می‌دهد که یک گروه انتهایی از یک متغیر انتخاب و سپس متغیر دیگری اندازه‌گیری شود؛ یعنی میانگین متغیر دوم در این گروه به میانگین همان متغیر در جمعیت نزدیک خواهد بود. به هر حال باید توجه داشت که این پدیده در سطح جمعیت رخ می‌دهد و در برای برخی از افراد مطالعه اندازه متغیر در اندازه‌گیری دوم ممکن است تفاوتی نکند یا بیش‌تر شود (۱۰).

^۱ Bland

^۲ Altman

^۳ Regression to mediocrity

با عملیات جبری ساده می‌توان نشان داد (۱۷) که مقدار منتظره بازگشت به میانگین از رابطه ۳ برابر است با:

$$C = \frac{\sigma_{\bar{X}}^2}{\sqrt{\sigma_{\bar{X}}^2 + \sigma_{\bar{Y}}^2}} \quad (۴)$$

در این رابطه $\sigma_{\bar{X}}^2$ واریانس خطای درون فردی است. اگر برای انتخاب و طبقه‌بندی افراد مطالعه n اندازه‌گیری تکراری انجام شود رابطه ۴ به صورت زیر تغییر می‌یابد (۱۷):

$$C = \frac{\sigma_{\bar{X}}^2/n}{\sqrt{\sigma_{\bar{X}}^2/n + \sigma_{\bar{Y}}^2/n}} \quad (۵)$$

این رابطه نشان می‌دهد که افزایش تعداد دفعات اندازه‌گیری برای انتخاب افراد اثر بازگشت به میانگین را کاهش می‌دهد (۱۷)، اما آن را حذف نمی‌کند. زیرا انحراف معیار خطا با توجه به ریشه دوم تعداد تکرارها تغییر می‌کند و در نتیجه تعداد اندازه‌گیری مورد نیاز برای حذف اثر بازگشت به میانگین خیلی زیاد خواهد بود (۱۹). مطالعه‌ها نشان داده که شیب کاهش در اثر بازگشت به میانگین تا ۴ اندازه‌گیری تند است، اما پس از آن افزایش تعداد اندازه‌گیری‌ها تأثیر کمی روی کاهش این اثر دارد (۱۳). هم‌چنین از فرمول شماره ۳ به راحتی می‌توان دریافت که مقدار منتظره بازگشت به میانگین به همبستگی بین دو اندازه‌گیری، مقدار آستانه برای انتخاب افراد و درجه تغییرپذیری در اندازه‌گیری‌ها بستگی دارد. هرچه همبستگی بین مقادیر اندازه‌گیری اول و دوم بیشتر باشد اثر بازگشت به میانگین کم‌تر خواهد بود و برعکس. هم‌چنین هرچه معیار انتخاب افراد انتهایی‌تر باشد اثر بازگشت به میانگین بیشتر خواهد بود. هرچه واریانس متغیر بزرگ‌تر باشد، سوءطبقه‌بندی افراد به گروه درمانی افزایش یافته و اثر بازگشت به میانگین بزرگ‌تر خواهد شد (۱۳).

در روش بیان شده در بالا فرض بر این است که خطاهای اندازه‌گیری و مقادیر واقعی متغیر مستقل از هم هستند، اما ممکن است این فرض همیشه برقرار نباشد. سهمی از خطای اندازه‌گیری مربوط به نوسانات بیولوژیک است که این سهم از خطا در اندازه‌گیری‌های تکراری ممکن است کاملاً مستقل از هم نباشند؛ برای مثال حالت‌های روحی و خلقی افراد روی اندازه فشار خون آن‌ها مؤثر است و در اندازه‌گیری‌های تکراری از فشار خون حالات روحی افراد ممکن است سبب گردد که نوسانات بیولوژیک در فشار خون کاملاً مستقل از هم نباشند، در این صورت محاسبه اثر بازگشت به میانگین پیچیده‌تر خواهد شد. جانسون و جورج (۱۷)

مستقل از یکدیگر باشند؛ ۲- مقدار خطا مستقل از مقدار واقعی متغیر باشد؛ ۳- توزیع خطاها و توزیع مقادیر واقعی در جمعیت نرمال باشد (که در نتیجه توزیع مقادیر مشاهده شده نیز نرمال خواهد بود)؛ و ۴- واریانس خطای درون فردی صرف‌نظر از مقدار واقعی متغیر برای همه افراد یکسان باشد (۱۶). برای محاسبه مقدار بازگشت به میانگین ابتدا میانگین منتظره متغیر در اندازه‌گیری اول از رابطه زیر محاسبه می‌شود (۱۳، ۱۶، ۱۷):

$$\mu + C\sigma = \text{میانگین منتظره در اندازه‌گیری اول} \quad (۱)$$

در این جا C آماره‌ای است که از رابطه زیر به دست می‌آید. یعنی:

$$C = \frac{\varphi(z)}{\Phi(z)} = \frac{1}{\sqrt{2\pi}} \frac{\exp(-\frac{1}{2}z^2)}{\int_z^\infty \varphi(u)du}$$

که Z برابر است با $\frac{(k-\mu)}{\sigma}$ و k حد آستانه برای انتخاب افراد است. جدول شماره ۲ که از مقاله یودکین و استارتون (۱۸) برگرفته شده، مقدار C را به ازای مقادیر مختلف z نشان می‌دهد. میانگین منتظره برای اندازه‌گیری دوم در این گروه به مقدار همبستگی بین دو اندازه در جمعیت بستگی دارد و برابر خواهد بود با:

$$\mu + C\sigma\rho = \text{میانگین منتظره برای اندازه‌گیری دوم} \quad (۲)$$

اگر مقادیر میانگین‌های منتظره برای دو اندازه‌گیری از هم کم شود، مقدار منتظره بازگشت به میانگین برابر خواهد بود با:

$$C\sigma(1-\rho) = \text{مقدار منتظره بازگشت به میانگین} \quad (۳)$$

اگر حد آستانه از میانگین جمعیت کم‌تر بوده و هدف انتخاب افرادی باشد که مقدار متغیر در آن‌ها کم‌تر از آستانه باشد؛ در فرمول‌های ۱ و ۲ به جای علامت به اضافه، علامت منها قرار می‌گیرد.

جدول شماره ۲- مقدار C برای مقادیر مختلف Z

z	0.00	0.02	0.04	0.06	0.08	0.10	0.12	0.14	0.16	0.18
0.00	0.80	0.81	0.82	0.84	0.85	0.86	0.88	0.89	0.90	0.92
0.20	0.93	0.94	0.96	0.97	0.98	1.00	1.01	1.03	1.04	1.05
0.40	1.07	1.08	1.10	1.11	1.13	1.14	1.16	1.17	1.19	1.20
0.60	1.22	1.23	1.25	1.26	1.28	1.29	1.31	1.32	1.34	1.35
0.80	1.37	1.38	1.40	1.41	1.43	1.45	1.46	1.48	1.49	1.51
1.00	1.53	1.54	1.56	1.57	1.59	1.61	1.62	1.64	1.65	1.67
1.20	1.69	1.70	1.72	1.74	1.75	1.77	1.79	1.80	1.82	1.84
1.40	1.85	1.87	1.89	1.90	1.92	1.94	1.96	1.97	1.99	2.01
1.60	2.02	2.04	2.06	2.08	2.09	2.11	2.13	2.15	2.16	2.18
1.80	2.20	2.21	2.23	2.25	2.27	2.28	2.30	2.32	2.34	2.36
2.00	2.37	2.39	2.41	2.43	2.44	2.46	2.48	2.50	2.52	2.53
2.20	2.55	2.57	2.59	2.61	2.62	2.64	2.66	2.68	2.70	2.71
2.40	2.73	2.75	2.77	2.79	2.80	2.82	2.84	2.86	2.88	2.90
2.60	2.91	2.93	2.95	2.97	2.99	3.01	3.02	3.04	3.06	3.08
2.80	3.10	3.12	3.13	3.15	3.17	3.19	3.21	3.23	3.25	3.26
3.00	3.28	3.30	3.32	3.34	3.36	3.38	3.39	3.41	3.43	3.45
3.20	3.47	3.49	3.51	3.53	3.54	3.56	3.58	3.60	3.62	3.64
3.40	3.66	3.68	3.69	3.71	3.73	3.75	3.77	3.79	3.81	3.83
3.60	3.85	3.86	3.88	3.90	3.92	3.94	3.96	3.98	4.00	4.02
3.80	4.04	4.05	4.07	4.09	4.11	4.13	4.15	4.17	4.19	4.21
4.00	4.23									

منبع: یودکین و استارتون (۱۸). مقادیر سطر و ستون اول نشان دهنده مقادیر Z و اعداد درون جدول مقدار C مربوطه را نشان می‌دهد.

با ۱/۶۶ بنابراین:

$$۵/۰۱ = ۴/۷۸۲ + ۱/۶۶ \times ۰/۱۳۷ = \text{میانگین منتظره برای}$$

اندازه‌گیری اول

$$۴/۹۴ = ۴/۷۸۲ + ۱/۶۶ \times ۰/۱۳۷ \times ۰/۶۸۶ = \text{میانگین منتظره}$$

برای اندازه‌گیری دوم

$$۰/۰۷ = ۴/۹۴ - ۵/۰۱ = \text{اثر منتظره بازگشت به میانگین}$$

ملاحظه می‌کنید که اثر منتظره بازگشت به میانگین ۰/۰۷ به‌دست آمده که با اثر مشاهده شده ۰/۰۶ = ۴/۹۶۸ - ۵/۰۲۸ تفاوت اندکی دارد، که می‌تواند مربوط به توزیع مقادیر فشارخون سیستولی باشد که با وجود استفاده از تبدیل لگاریتمی کاملاً نرمال نشده است. بنابراین می‌توان نتیجه گرفت که تفاوت مشاهده شده در فشار خون سیستولی این افراد در فازهای اول و دوم ناشی از پدیده بازگشت به میانگین است.

جدول شماره ۳- میانگین و انحراف معیار فشارخون سیستولی و تبدیل لگاریتمی آن را برای مردان ۳۰-۷۴ ساله در فازهای اول و دوم مطالعه قند و لیپید تهران

کل مردان (n = ۱۳۴۷)	افراد با فشارخون بالای ۱۴۰ میلی‌متر جیوه (n = ۱۷۳)		
	میانگین (انحراف)	میانگین (انحراف)	میانگین برای تبدیل لگاریتمی (انحراف معیار)
فاز اول	۱۲۰/۵ (۱۷/۲)	۱۵۴/۲ (۱۳/۰)	۵/۰۲۸ (۰/۰۸۱)
فاز دوم	۱۲۰/۳ (۱۷/۸)	۱۴۵/۰ (۱۹/۳)	۴/۹۶۸ (۰/۱۳۱)
همبستگی دو اندازه	۰/۶۸۷	۰/۶۸۶	-

بازگشت به میانگین در مطالعه‌های اپیدمیولوژی و تأثیر آن در نتیجه‌گیری و تفسیر نتایج

موارد متعددی از رخداد این پدیده در آموزش و تحصیلات (۲۶)، ورزش (۲۷) و سایر رشته‌ها یاد شده است. در زمینه علوم پزشکی و بهداشتی رخداد این پدیده در مطالعه‌های مختلف نشان داده شده است. بلاند و آلتمن (۱۱) ۵ مثال از رخداد این پدیده را بیان کرده‌اند. دیویس (۱۳) در مقاله ارزشمند خود دو گروه از مطالعاتی را که بازگشت به میانگین در آن‌ها رخ می‌دهد؛ بیان

در مقاله خود این موضوع را مورد بررسی قرار داده و فرمول‌های مربوطه را ارائه کرده‌اند. علاوه بر این در طی سال‌های اخیر پژوهش در این زمینه فعال بوده و مقاله‌های جدید در مورد روش‌های متعدد پارامتری و ناپارامتری برای برآورد اثر بازگشت به میانگین برای توزیع‌های نرمال و غیر نرمال چاپ شده است (۲۵-۲۰). می^۱ و چوا^۲ روشی را مبتنی بر آزمون t زوجی ارائه کرده‌اند که می‌تواند سهم مربوط به درمان را از بازگشت به میانگین جدا کند، اما استفاده از این روش مبتنی بر این است که اطلاعات مربوط به میانگین جمعیت در دسترس باشد (۲۵).

مثالی از اثر بازگشت به میانگین

در این جا یک مثال واقعی از این پدیده از داده‌های مطالعه قند و لیپید تهران ذکر می‌شود. این مطالعه از سال ۱۳۷۸ با هدف برآورد میزان شیوع و بروز اختلالات متابولیک و عوامل خطر بیماری‌های مهم غیر واگیر در حال اجرا است. برای مثال حاضر از داده‌های ۱۳۴۷ مرد ۳۰-۷۴ ساله که در فازهای اول تا سوم مطالعه شرکت داشتند استفاده شد. این افراد در شروع مطالعه سابقه بیماری قلبی-عروقی نداشته و هم‌چنین هیچ‌کدام از آن‌ها داروی ضد فشارخون مصرف نمی‌کردند. ۱۷۳ نفر از این مردان در شروع مطالعه فشار خون سیستولی بیش‌تر از ۱۴۰ میلی‌متر جیوه داشتند، که ۳ سال بعد در هنگام دومین اندازه‌گیری میانگین فشارخون این افراد ۸ میلی‌متر جیوه کاهش را نشان داد. آیا این تفاوت ناشی از بازگشت به میانگین بوده یا این که واقعاً فشارخون سیستولی آن‌ها به دلیل مثلاً تغییر در شیوه زندگی، کاهش داشته است؟ در این جا با استفاده از فرمول‌های معرفی شده در بخش قبل بررسی می‌شود که آیا این تفاوت‌ها ناشی از بازگشت به میانگین است یا خیر؟ در مثال حاضر از آن جا که توزیع فشارخون دارای چولگی به سمت راست بود، از تبدیل لگاریتم طبیعی مقادیر برای محاسبات استفاده شد. جدول شماره (۳) میانگین و انحراف معیار فشار خون سیستولی و تبدیل لگاریتمی آن را برای افراد یاد شده در فازهای اول و دوم مطالعه نشان می‌دهد. تبدیل لگاریتمی حد آستانه انتخابی $\ln(۱۴۰) = ۴/۹۴۲$ بود؛ بنابراین:

$$z = \frac{۴/۹۴۲ - ۴/۷۸۲}{۰/۱۳۷} = ۱/۱۷$$

با استفاده از جدول شماره ۲ مقدار C برای $z = ۱/۱۷$ برابر است

^۱ Mee

^۲ Chua

مانند کارآزمایی‌های بالینی و پیشگیری و مطالعات ارزیابی برنامه‌ها مورد بررسی قرار می‌گیرند که در آن‌ها به منظور پایش اثر مداخله یا برنامه، اندازه‌گیری‌های تکراری از متغیر برای افراد انجام می‌شود. در کارآزمایی‌های بالینی در موارد زیر این پدیده رخ خواهد داد:

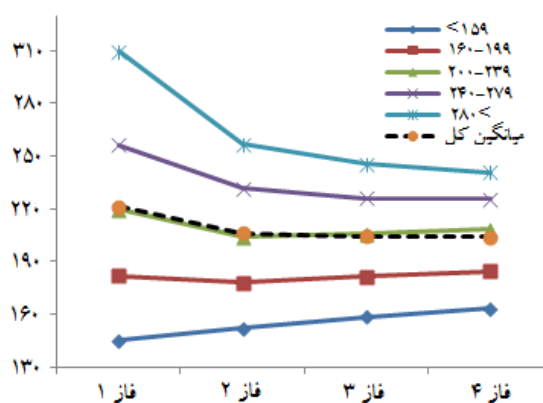
اول در مطالعه‌هایی که در آن‌ها افراد در طی دو مرحله برای شرکت در مطالعات انتخاب می‌شوند یعنی ابتدا افرادی که سطح متغیر مورد نظر در آن‌ها از حد معینی بیشتر (یا کمتر) باشد؛ نامزد شرکت در مطالعه می‌شوند. سپس به فاصله زمانی کوتاهی این افراد مجدداً برای بررسی واجد شرایط بودن مورد معاینه مجدد قرار می‌گیرند. در این مرحله تعدادی از نامزدهای انتخابی به دلیل بازگشت به میانگین واجد شرایط شرکت در کارآزمایی نخواهند بود. به این ترتیب اثر بازگشت به میانگین می‌تواند روی شرکت افراد در کارآزمایی‌های بالینی اثر گذاشته و در اجرای کارآزمایی مشکل ایجاد کند (۳۰، ۱۳).

دوم در کارآزمایی‌های بالینی که در آن‌ها هیچ گروه شاهدی برای مقایسه در نظر گرفته نمی‌شود، مانند کارآزمایی‌های قبل و بعد (۱۳). مانند مطالعه‌هایی که هدف آن‌ها ارزیابی اثر یک روش درمانی یا پیشگیری در کاهش (یا افزایش) مقادیر یک متغیر مانند فشارخون (یا HDL کلسترول) است. در این مطالعه‌ها حتی اگر روش درمانی یا پیشگیری هیچ تأثیری نداشته باشد؛ مقدار میانگین متغیر افراد شرکت کننده در مطالعه در پایش اثر درمانی به دلیل بازگشت به میانگین تغییر خواهد کرد و این طور به نظر خواهد رسید که درمان مؤثر بوده در حالی که هیچ اثری نداشته است، یا حتی اگر درمان مؤثر هم بوده باشد؛ اثر درمانی بیش‌تری در گروه مورد مطالعه مشاهده خواهد شد. این موضوع منحصر به کارآزمایی‌های بالینی نیست، بازگشت به میانگین ارزیابی بیش‌تر برنامه‌های پیشگیری و درمانی مبتنی بر اندازه‌گیری‌های پیش و پس را که روی جمعیت‌هایی که بر اساس مقدار قبل از برنامه انتخاب شده‌اند، مخدوش می‌کند (۳۱).

سوم در کارآزمایی‌های شاهددار که در آن‌ها تخصیص افراد به گروه‌های درمان و کنترل با تصادفی سازی انجام نمی‌گیرد. در این مطالعه‌ها اگر نسبت بالایی از افرادی که به گروه درمان تخصیص می‌یابند؛ از بیمارانی باشند که بیماری آن‌ها شدیدتر است در ارزیابی دوم و پایش اثرات درمانی این بیماران این افراد به دلیل بازگشت به میانگین اثر درمانی بیش‌تری را در مقایسه با گروه دیگر نشان خواهند داد و این طور به نظر خواهد رسید که درمان

می‌کند. مورتون و تورگرسون (۲۸، ۲۹) نیز رخداد این پدیده در ۶ موقعیت بیان کرده‌اند.

اثر بازگشت به میانگین در مطالعه‌هایی قابل مشاهده است که در آن اندازه‌گیری‌های تکراری برای یک گروه انتخابی و غیرتصادفی از افراد وجود داشته باشد. بنابراین در مطالعه‌های اکولوژیک، مورد-شاهدی و مقطعی اثر بازگشت به میانگین مشاهده نخواهد شد. در مطالعه‌های هم‌گروهی معمولاً متغیرهای مورد بررسی در فواصل زمانی معینی اندازه‌گیری می‌شوند، اما همگروه این مطالعات را افرادی از کل طیف توزیع متغیر تشکیل می‌دهند و با وجود این که در اندازه‌گیری‌های تکراری مقادیر به‌دست آمده برای افراد، به ویژه آن‌هایی که در دو انتهای توزیع قرار دارند با مقدار قبلی متفاوت است، میانگین‌های به‌دست آمده تفاوت معنی‌داری نخواهد داشت. اگر در این مطالعه‌ها افراد مورد بررسی را بر حسب نخستین اندازه‌گیری از متغیر به چند گروه تقسیم کرده و برای افراد هر گروه مقدار میانگین در اندازه‌گیری اول و اندازه‌گیری‌های بعدی به‌دست آورده شود؛ این پدیده در گروه‌های انتهایی توزیع قابل مشاهده خواهد بود، اما میانگین کل همگروه، تنها به دلیل وجود روندهای افزایشی یا کاهش‌ی متغیر در طول زمان تغییر خواهد کرد. برای مثال به شکل شماره ۱ توجه کنید. این شکل مربوط به میانگین کلسترول زنان ۲۴-۳۰ ساله در مطالعه همگروهی قند و لیپید تهران است، خط نقطه چین، میانگین کلسترول را برای کل همگروه در فازهای اول تا چهارم و خطوط توپر برای گروه‌های کلسترول خون بر اساس نخستین اندازه‌گیری از متغیر نشان می‌دهد.



شکل شماره ۱- میانگین کلسترول زنان ۲۴-۳۰ ساله در ۴ فاز مطالعه همگروهی قند و لیپید تهران
گروه‌های انتخابی از جمعیت عموماً در مطالعه‌های مداخله‌ای

مؤثرتر بوده است.

چهارم در مطالعه‌هایی که در آن‌ها به گروه شاهد دارونما داده می‌شود، تا بیمار از عدم دریافت مداخله بی‌اطلاع باشند (کورسازی). گاهی افراد دریافت‌کننده دارونما، پاسخ‌های قوی را نشان می‌دهند. این موضوع به ویژه در بین بیماران دارای علائم شدید بیماری، و به ویژه در مورد متغیرهای ذهنی^۱ (که در طی زمان ماهیت متغیری دارند و در نتیجه احتمال خطا در سنجش و اندازه‌گیری آن‌ها زیاد است) مانند شدت درد، علائم یائسگی یا علائم افسردگی بارزتر است. بخشی از پاسخ به دارونما در این مطالعه‌ها می‌تواند به علت بهبود خودبه‌خودی یا نوسانات طبیعی در بیماری (بازگشت به میانگین) است.

با توجه به آن‌چه که در مورد اثر این پدیده روی نتیجه‌گیری و تفسیر یافته‌های مطالعه‌های مداخله‌ای بیان شد بازگشت به میانگین تهدیدی برای اعتبار داخلی مطالعه‌ها محسوب می‌شود (۳۲).

اجتناب از بازگشت به میانگین

از آن‌جا که بازگشت به میانگین با طراحی و اجرای کارآزمایی‌های بالینی ارتباط تنگاتنگی دارد، مهم‌ترین اقدام در نظر گرفتن این پدیده به هنگام طراحی مطالعه است. وجود یک گروه شاهد همزمان و تخصیص تصادفی افراد به گروه‌ها به منظور کنترل یا حذف همه منابع تورش شناخته شده و شناخته نشده به طور برابر بین گروه‌های کنترل و درمان سبب خواهد شد که دو گروه از این نظر کاملاً مشابه هم باشند. در نتیجه بازگشت به میانگین در دو گروه یکسان رخ داده و هر تفاوتی بین دو گروه، مربوط به اثر درمان خواهد بود (۱۳،۳۳). در مطالعه‌های شاهددار، به ویژه در مطالعه‌هایی که افراد دارای شرایط از طریق غربالگری وارد مطالعه می‌شوند، یک مشکل دیگر انتخاب یک اندازه پایه برای اندازه‌گیری اثر درمانی است. یک راه حل پیشنهادی این است که ابتدا یک اندازه‌گیری برای طبقه‌بندی افراد انجام شود. اندازه‌گیری دوم به عنوان اندازه پایه متغیر برای افراد منظور شده و پس از درمان اندازه‌گیری سوم برای مقایسه انجام شود (۳۴).

راه حل دیگر کاهش خطای تصادفی در اندازه‌گیری با انجام چندین اندازه‌گیری و گرفتن میانگین از اندازه‌ها است (۱۶). برای این کار یا باید تعداد دفعات اندازه‌گیری در یک معاینه افزایش یابد، یا این که چندین اندازه‌گیری در فواصل زمانی معین انجام شود. افزایش تعداد اندازه‌گیری‌ها در یک نوبت معاینه، اثر خطاهای

ناشی از تکنیک اندازه‌گیری را کاهش می‌دهد. افزایش تعداد اندازه‌گیری‌ها در زمان‌های مختلف می‌تواند هم اثر خطای ناشی از تغییر درون فردی و هم خطای تکنیکی را کاهش دهد، اما باید توجه داشت که با افزایش فاصله زمانی بین اندازه‌گیری‌ها ممکن است الگوی همبستگی بین اندازه‌گیری‌ها تغییر کند. معمولاً با افزایش فاصله بین اندازه‌گیری‌ها همبستگی بین دو اندازه‌گیری کاهش می‌یابد (۱۷). به هر حال افزایش تعداد اندازه‌گیری‌ها به نوع متغیر مورد بررسی، هزینه و زمان مورد نیاز برای انجام مطالعه وابسته است. در مورد اندازه‌گیری‌های غیر تهاجمی مانند اندازه فشار خون می‌توان چندین بار اندازه‌گیری انجام داد، اما این کار مستلزم صرف زمان است. در مورد اندازه‌گیری‌های تهاجمی مانند آن‌هایی که به نمونه خون یا بافت‌های بدن نیاز دارند، این کار ممکن است از نظر زمان و هزینه و اخلاق پژوهش امکان‌پذیر نباشد. به هر حال همان‌طور که قبلاً نیز گفته شد این روش اگر چه بازگشت به میانگین را کاهش می‌دهد، اما آن را کاملاً حذف نمی‌کند.

یک راه دیگر برآورد مقدار اثر بازگشت به میانگین در مطالعه به یکی از روش‌های بیان شده در بالا است. به هر حال در هنگام اجرای مطالعه‌های بالینی ممکن است برآوردهای مربوط به میانگین و انحراف معیار جمعیت و هم‌چنین همبستگی بین دو اندازه‌گیری ممکن است موجود نباشد، در این صورت می‌توان برآوردهای این پارامترها از مطالعه‌های مرتبط و مشابه به دست آورد. در صورتی که برآوردهای مورد نیاز از دیگر مطالعه‌ها در دسترس نباشد، جیمز (۳۵) روشی را معرفی کرده که می‌توان مقادیر این پارامترها را از داده‌های موجود به دست آورد و محاسبه‌ها را انجام داد.

و در نهایت در مرحله تجزیه و تحلیل داده‌ها می‌توان از آنالیز کوواریانس برای مقایسه اثرات درمانی بین دو گروه مورد مطالعه استفاده کرد، که در آن اندازه‌های پایه به عنوان یک متغیر کمکی برای کنترل منابع تغییر در افراد مورد استفاده قرار خواهد گرفت (۱۳،۳۶).

نتیجه‌گیری

در این مقاله به پدیده «بازگشت به میانگین» به عنوان یکی از اثرات خطا در اندازه‌گیری متغیرها، علت رخداد این پدیده و این که در چه نوع مطالعه‌هایی رخ می‌دهد؛ پرداخته شد. نکته مهم توجه به این مسأله است که بازگشت به میانگین تقریباً در همه جا وجود دارد و می‌تواند روی نتیجه‌گیری و تفسیر یافته‌های

^۱ Subjective

کارآزمایی بالینی می‌توانند با قرار دادن این موضوعها در سرفصل‌های برنامه‌های کارگاه‌های آموزشی نقش مهمی را در آشنایی بیشتر پژوهشگران بالینی کشور با موضوع بازگشت به میانگین و اثرهای ناشی از آن داشته باشد.

تشکر و قدردانی

این مقاله بر گرفته از پایان نامه خانم سیما مسعودی برای اخذ درجه دکتری اپیدمیولوژی از دانشکده بهداشت دانشگاه علوم پزشکی شهید بهشتی می باشد.

مطالعه‌ها به ویژه مطالعه‌های مداخله‌ای و مطالعه‌هایی که تأثیر برنامه‌های پیشگیری و درمانی در آنها ارزیابی می‌گردد، اثر بگذارد. بنابراین در طراحی، اجرا و آنالیز مطالعه‌ها باید تدبیرهای لازم برای اجتناب از اثرات آن اتخاذ گردد. در نهایت برای مشاهده اثرات این پدیده در مطالعه‌ها، روش‌هایی وجود دارد که با آنها می‌توان اثر بازگشت به میانگین را از اثر مداخله تمیز داد.

با وجود اهمیتی که خطای تصادفی در اندازه‌گیری متغیرها و پدیده بازگشت به میانگین روی نتیجه‌گیری و تفسیر نتایج مطالعه‌ها دارند، این موضوع در بیشتر مطالعه‌ها نادیده گرفته می‌شوند. در این بین برنامه‌ریزان و مدرسان کارگاه‌های آموزشی طراحی و اجرای مطالعه‌ها اپیدمیولوژیک و به ویژه مطالعه‌های

منابع

- Gustafson P. Measurement error and misclassification in statistics and epidemiology: Impacts and bayesian adjustments. Florida, USA: Chapman & Hall/CRC; 2004: 8.
- Bland JM, Altman DG. Measurement error. *BMJ*. 1996 Sep 21; 313: 744.
- Crocker L, Aigina J. Introduction to Classical and Modern Test Theory. Ohio, USA: Cengage Learning; 2008:110.
- Rabinovich SG. Measurement Errors and Uncertainties: Theory and Practice. 3rd edition. NJ USA: Springer; 2005:1-28.
- Last JM. A dictionary of epidemiology. 4th edition. New York: Oxford University Press; 2001: 63.
- Zhang X, Tomblin JB. Explaining and Controlling Regression to the Mean in Longitudinal Research Designs. *J o Juorunranla ol fo Sf pSepeecehc,h L, alnagnugaugaega,e a, nahn dH Heaerainrign gR eRseesaeracrh*. 2003; 46: 1340-51.
- Kirkwood BR, Sterne JAC. Essential medical statistics. Malden, Massachusetts, USA: Blackwell; 2003: 444.
- Scott M. Grundy, James I. Cleeman, C. Noel Bairey Merz, H. Bryan Brewer J, Luther T. Clark, Donald B. Hunninghake, et al. Implications of Recent Clinical Trials for the National Cholesterol Education Program Adult Treatment Panel III Guidelines. *Circulation*. 2004; 110: 227-39.
- Chobanian AV, Bakris GL, Black HR, Cushman WC, Green LA, J L Jr. Izzo JL, et al. The seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: The JNC 7 Report. *JAMA*. 2003; 289: 2560-72.
- Streiner DL. Regression Toward the Mean: Its Etiology, Diagnosis, and Treatment. *Research Methods In Psyciatry*. 2201; 46: 72-6.
- Bland JM, Altman DG. Some examples of regression towards the mean. *British Medical Journal*. 1994;309:780.
- Bland JM, Altman DG. Regression towards the mean. *British Medical Journal*. 1994; 308(6942):1499.
- Davis CE. The effect of regression to the mean in epidemiologic and clinical studies. *American Journal of Epidemiology*. 1976; 104: 493-8.
- Fitzmaurice G. Measurement error and reliability. *Nutrition*. 2002; 18: 112-4.
- Galton F. Regression Towards Mediocrity in Hereditary Stature. *The Journal of the Anthropological Institute of Great Britain and Ireland*. 1886; 15: 246-63.
- Gardner MJ, Heady JA. Some effects of within-person variability in epidemiological studies. *J Chron Dis* 1973; 26: 781 -95.
- Johnson WD, George VT. Effect of regression to the mean in the presence of within-subject variability. *Statistics in Medicine*. 1991; 10: 1295-302.
- Yudkin PL, Stratton IM. How to deal with regression to the mean in intervention studies. *Lancet*. 1996; 347: 241-3.
- Blomqvst N. On the bias caused by regression toward the mean in studying the relation between change and initial value. *Journal of Clinical Periodontology*. 1986; 13: 34-7.
- Lin HM, Hughes MD. Adjusting for regression toward the mean when variables are normally distributed. *Statistical Methods in Medical Research*. 1997; 6: 129-46.
- Chesher A. Non-normal variation and regression to the mean (abstract). *Statistical Methods in Medical Research*. 1997; 6: 147-66.
- Muller H-G, Abramson I, Azari R. Nonparametric regression to the mean. *Proceedings of the National Academy of Sciences of the United States of America* 2003; 100: 9715-20.
- Ostermann T, Willich SN, Lüdtke R. Regression toward the mean – a detection method for unknown population mean based on Mee and Chua's algorithm. *BMC Medical Research Methodology*. 2008; 8: 52.
- Kowalskia CJ, Schneiderman ED, Willis SM. PC program implementing an alternative to the paired t-test which adjusts for regression to the mean. *International Journal of Bio-Medical Computing* 1994; 37: 189-94.
- Mee RT, Chua TC. Regression toward the mean and the paired sple test (abstract). *American Statistition*. 1991; 45: 39-42.
- Smith G, Smith J. Regression to the Mean in Average Test Scores. *Educational Assessment*. 2005; 10(4): 377-399
- Shephard RJ. Regression to the Mean A Threat to Exercise Science? *Sports Medicine*. 2003; 33: 575-84.
- Morton V, Torgerson DJ. Effect of regression to the mean on decision making in health care. *British Medical Journal*. 2003; 326: 1083-4.
- Morton V, Torgerson DJ. Regression to the mean: treatment

- effect without the intervention. *Journal of Evaluation in Clinical Practice*. 2005; 11: 59-65.
30. Davis CE. Regression to the Mean. *Wiley Encyclopedia of Clinical Trials*: John Wiley & Sons, Inc.; 2007.
 31. Shepard DS, Finison LJ. Blood pressure reductions: correcting for regression to the mean. *Preventive Medicine*. 1983; 12: 304-17.
 32. Smith RA. Regression to the mean. In: Everitt BS, Howell DC, editors. *Encyclopedia of Statistics in Behavioral Science*. Chichester: John Wiley & Sons, Ltd; 2005. 731–1732.
 33. Linden A. Estimating the effect of regression to the mean in health management programs. *Disease Management & Health Outcomes*. 2207; 15: 7-12.
 34. Ederer F. Serum cholesterol changes: effects of diet and regression toward the mean. *Journal of Chronic Disease*. 1972; 25: 277-89.
 35. James KE. Regression toward the mean in uncontrolled clinical studies. *Biometrics*. 1973; 29: 121-30.
 36. Vickers AJ, Altman DG. Statistics notes: analyzing controlled trials with baseline and follow up measures. *BMJ*. 2001; 323: 1123–4.

Regression to the Mean in Epidemiologic Studies

Masudi S¹, Mehrabi Y², Khalili D³, Yavari P⁴

1- Urmia Medical Sciences University, Urmia, Iran, Department of Epidemiology, School of Public Health, Shahid Beheshti University of Medical Sciences, Tehran, Iran

2- Department of Epidemiology, School of Public Health, Shahid Beheshti University of Medical Sciences, Tehran, Iran

3- Prevention of Metabolic Disorders Research Center, Research Institute for Endocrine Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran

4- Department of Health and Community Medicine, School of Medicine, Shahid Beheshti University of Medical Sciences and Iranian Epidemiological Association, Tehran, Iran

Corresponding author: p.yavari-grc@sbmu.ac.ir

In epidemiologic studies, the measurement of characteristics of interest is almost always subject to random measurement error. This error and its effects are usually overlooked by researchers. One of its effects is a widespread statistical phenomenon that is well known as regression to the mean. This phenomenon occurs whenever an extreme group of people is selected from a population based on their measurements of a variable. If a second measurement is taken in this group, the mean of the second measurement will be closer to the mean of the population. In interventional studies, this increase (decrease) might be regarded as the effect of intervention, when in fact it has had no effect. Ignoring regression to the mean will lead to the erroneous conclusions and interpretation of the results of epidemiologic studies and affects the decisions in evidence-based medicine and planning for preventive and public health measures. This paper highlights the importance of this problem and its effects in epidemiologic studies and the ways to avoid it.

Key Words: Random measurement error, Regression dilution, Epidemiologic study