مجله‌تخصصی اپیدمیولوژی ایران؛ ۱۳۹۰، دوره ۷، شماره ۲: صفحات ۷۴-۶۷.
مقاله مروری

چگونه حجم نمونه را در شرایط خاص تخمین بزنیم ؟

علی اکبر حق دوست[۱]، محمد رضا بانشی[۲]، مریم مرزبان[۴و۳]

[۱] دانشیار اپیدمیولوژی، مرکز تحقیقات مدل سازی در سلامت، دانشگاه علوم پزشکی کرمان، ایران
[۲] استادیار آمارزیستی، مرکز تحقیقات فیزیولوژی، دانشگاه علوم پزشکی کرمان، ایران
[۳] مرکز تحقیقات طب سنتی و تاریخ طب دانشگاه علوم پزشکی شیراز، ایران
[۴] دانشجوی کارشناسی ارشد اپیدمیولوژی، کمیته تحقیقات دانشجویی، دانشگاه علوم پزشکی کرمان، ایران

نویسنده مسئول: مریم مرزبان، نشانی: شیراز، دانشگاه علوم پزشکی شیراز، مرکز تحقیقات طب سنتی و تاریخ طب، دانشکده پزشکی، طبقه هشتم، تلفن: ۲۳۳۷۵۸۹ -۰۷۱۱،
پست الکترونیک: marzbanh@gmail.com

تاریخ دریافت:۸۹/۶/۲۷؛ پذیرش: ۸۹/۱۱/۲

مقدمه و اهداف: در مقاله قبلی درباره پاره‌ای از مباحث کلیدی نمونه‌گیری در پژوهش بحث شد. در این مقاله سعی بر آن است که برای تعدیل حجم نمونه در شرایط خاص مانند مقایسه‌های چندگانه؛ محاسبه حجم نمونه برای مواردی که تعداد نمونه‌ها در گروه‌ها برابر نیست، تصحیح حجم نمونه برای مقادیرنامشخص و تعدیل حجم نمونه برای جامعه محدود مطالبی ارایه شود. به علاوه، درباره اثر طرح در نمونه‌گیری چند مرحله‌ای و تأثیر آن در حجم نمونه مطالبی بیان شده است. سپس به تخمین حجم نمونه هنگامی که آزمون‌های ناپارامتریک برای آنالیز داده‌ها به کار برده می‌شود پرداخته شود. همچنین توضیحاتی در مورد توان کارآمدی شیوه‌های پارامتریک در برابر غیر پارامتریک و کاربردشان در تصحیح حجم نمونه ارایه گردد.

واژگان کلیدی: حجم نمونه مقایسه‌های چندگانه، اثر طرح، آزمون‌های ناپارامتری، مقادیر گمشده، جامعه محدود

مقدمه

همانطور که درقسمت قبل بیان شد مبحث حجم نمونه و نحوه محاسبه آن یکی از مباحث کلیدی در تحقیق بوده و برای محاسبه آن باید بسته به نوع متغیرها و نوع مطالعه از فرمول‌های خاصی با پیش فرض‌های منطقی استفاده نماییم که در مقاله قبلی به تفصیل در این باره مطالبی ارائه شد (۱). همچنین به مفاهیم پایه مانند میزان اثر، دامنه و ضریب اطمینان، و خطاهای آماری و مفروضات مورد نیاز برای محاسبه حجم نمونه اشاره و شیوه محاسبه حجم نمونه در ساده‌ترین شکل ممکن شرح داده شد.

در این مقاله و در ادامه مباحث قبلی به زوایای پیچیده‌تری از محاسبه حجم نمونه درشرایط خاص اشاره خواهد شد. موارد مهمی که در این مقاله به آن‌ها پرداخته خواهد شد عبارت هستند از :

۱. برآورد حجم نمونه برای مقایسه‌های چندگانه
۲. برآورد حجم نمونه در مطالعات unbalance
۳. تصحیح حجم نمونه برای مقادیرنامشخص
٤. برآورد حجم نمونه برای روش نمونه گیری چند مرحله‌ای (multistage sampling)
۵. برآورد حجم نمونه برای آزمون‌های ناپارامتری
٦. برآورد حجم نمونه برای جامعه محدود

برآورد حجم نمونه برای مقایسه‌های چندگانه

فرمول‌های ارایه شده برای محاسبه حجم نمونه در مقاله قبلی، برای زمانی هستند که شما قصد داشته باشید دو گروه را با یکدیگر مقایسه نمایید. ولی در تحقیقات ممکن است شما مجبور به مقایسه بیش از دو گروه با هم باشید، آیا در این موارد می‌توانیم به سادگی با همان حجم نمونه قبلی کار را بر روی سه گروه انجام دهیم؟ به عنوان مثال فرض کنید قصد دارید یک کارآزمایی بالینی برای مقایسه سه دارو طراحی کنید. بر اساس پیش فرض‌های ارایه شده و بر اساس فرمول مقایسه دو میانگین حجم نمونه برای هر گروه ۱۰۰ بدست آمده است. حال برای انجام این تحقیق آیا می‌توانید سه گروه با حجم نمونه ۱۰۰ نفر بگیریم و در انتها با همان دقت آماری تحلیل خود را انجام دهیم؟

متاسفانه جواب منفی است و باید بر اساس تعداد گروه‌های مورد مطالعه حجم نمونه تعدیل گردد. اگرچه فرمول‌های دقیق این تعدیل پیچیده است ولی خوشبختانه یک فرمول بسیار ساده برای زمانی که تعداد گروه‌های مورد مقایسه کمتر از ۵ گروه هستند وجود دارد. اگر تعداد گروه‌های مورد مقایسه را با g نشان دهیم، حجم نمونه تعدیل شده برای هر گروه برابر است با (۲).

$$n' = n * \sqrt{g-1}$$

به عنوان مثال اگر حجم نمونه اولیه ۱۰۰ نفر برای مقایسه دو دارو باشد، برای مقایسه سه دارو (g=۳) بایست مقدار ۱۰۰ را در جذر ۲ یعنی ۱/۴ ضرب نمود و در هر گروه به جای ۱۰۰ نفر، ۱۴۰ نفر وارد مطالعه نمود. به عبارتی حجم نمونه کل برابر ۳۰۰ نفر نخواهد بود بلکه بایست ۴۲۰ نفر وارد مطالعه شوند.

همانگونه که ملاحظه می‌فرمایید با افزایش تعداد گروه‌های مورد مقایسه، حجم نمونه هر گروه به صورت قابل ملاحظه‌ای افزایش می‌یابد و شاید یکی از دلایلی که عموماً توصیه می‌شود که تعداد گروههای مورد مقایسه به حداقل برسد، همین نکته باشد.

البته یک مورد حالت استثنا وجود دارد و آن زمانی است که یک گروه رفرنس وجود دارد و سایر گروه‌ها فقط با گروه رفرنس مقایسه می‌شوند و مقایسه بین سایر گروه‌ها اصلاً مد نظر نیست (۲).

به عنوان مثال فرض کنید قصد شما مقایسه اثربخشی دو داروی جدید در مقایسه با یک داروی استاندارد است و فقط می‌خواهید ببینید که آیا این داروها در مقایسه با داروی استاندارد تفاوت اثر دارند یا خیر و مقایسه بین اثربخشی این دو دارو جزو اهداف اصلی مطالعه شما نیست. در این گونه موارد بعضی معتقد هستند (۲) که با استفاده از فرمول فوق فقط باید حجم نمونه گروه رفرنس را تعدیل نمود و حجم نمونه سایر گروه‌ها را می‌توان برابر همان مقدار قبلی که از فرمول مقایسه دو گروه بدست می‌آید در نظر گرفت. با توجه به این توضیحات در مورد مثال قبل باید گروه مبنا را ۱۴۰ بگیریم و دو گروه دیگر را همان ۱۰۰ نفر در نظر داشته باشیم.

برآورد حجم نمونه در مطالعات unbalance

آنچه در فرمول‌های بیان شده قبلی بدست می‌آید با این شرط بوده که حجم نمونه در دو گروه برابر است و به عبارتی تعادل (balance) در حجم نمونه گروه‌های مورد مقایسه وجود داشته است. ولی درشرایط متعددی ممکن است حجم نمونه دو گروه برابر نباشد. به عنوان مثال فرض نمایید که در یک مطالعه تحلیلی قصد بر آن است که تأثیر مصرف سیگار بر بالا بودن فیبرینوژن خون (یک فاکتور خونی است که بالا بودن آن احتمال ایجاد لخته در عروق خونی را افزایش می‌دهد) در بیماران مبتلا به سکته قلبی بررسی کنیم. اگر حجم نمونه محاسبه شده برابر ۱۰۰ نفر باشد بدین معنا است که بایست ۱۰۰ نفر غیر سیگاری و ۱۰۰ نفر سیگاری وارد مطالعه شوند. ولی اگر نمونه‌گیری به صورت متوالی در اورژانس قلب صورت پذیرد، قطعاً چنین تناسب ۱ به ۱ برای افراد سیگاری و غیر سیگاری محقق نخواهد شد، مثلاً اگر تقریباً ۲۰٪ افراد مبتلا به سکته قلبی مراجعه کننده به اورژانس سیگاری باشند، بدین معنی است که به ازای ۱ سیگاری ما با ۴ غیر سیگاری برخورد خواهیم نمود و اگر به صورت متوالی نمونه‌گیری را شروع کنیم و با رسیدن به نفر ۲۰۰ام، نمونه‌گیری را متوقف نماییم، احتمالاً در این فرآیند ۴۰ سیگاری و ۱۶۰ غیر سیگاری وارد مطالعه خواهند شد. برای حل این مشکل یا باید به صورت عمد نمونه‌گیری را از حالت ساده و متوالی خارج ساخته و به گونه ای نمونه‌گیری کنیم که در انتها تعادل ۱ به ۱ در حجم نمونه گروه‌ها بوجود آید و یا باید با یک تصحیح مشخص سعی کنیم به اعدادی برای حجم نمونه افراد سیگاری و غیر سیگاری به نسبت ۱ به ۴ برسیم که از نظر آماری همان قدرت ۱۰۰ حجم نمونه با نسبت ۱ به ۱ را داشته باشد.

فرمول تصحیح حجم نمونه برای زمانی که نسبت حجم نمونه در دو گروه برابر نیست بسیار ساده است. ابتدا باید این نسبت را حدس بزنیم. برای سادگی کار از این به بعد نسبت حجم نمونه در دو گروه را با **C** نمایش می‌دهیم. بر اساس توضیحات ارایه شده در مثال فوق مقدار **C** برابر ۴ است و حجم نمونه بدست آمده از فرمول‌های قبلی با لحاظ نمودن تعادل در حجم نمونه‌ها (یعنی **C** برابر ۱) برابر ۱۰۰ نمونه در هر گروه بود.

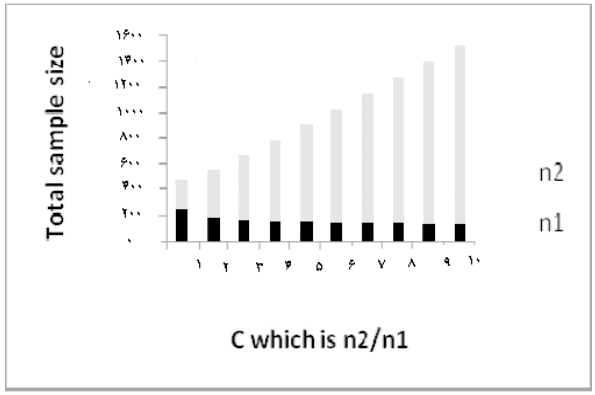
فرمول تصحیح بسیار ساده است. ابتدا شما با استفاده از فرمول زیر بایست حجم نمونه برای گروه کوچک‌تر (در این مثال یعنی افراد سیگاری) را حساب نمایید و بعد آن را در **C** ضرب نمایید تا حجم نمونه برای گروه بزرگتر (در این مثال یعنی افراد غیر سیگاری) بدست آید (۳)

$$n' = \frac{c+1}{2c} * n$$

بر اساس توضیحات فوق اگر مقدار **C** را برابر ۴ و مقدار **n** را برابر ۱۰۰ قرار دهید، حجم نمونه مورد نیاز برای گروه سیگاری تقریباً ۶۳ نفر بدست خواهد آمد و در مقابل حجم نمونه برای گروه غیرسیگاری ۴ برابر این مقدار یعنی ۲۵۲ نفر خواهد بود. به عبارتی در کل ما به جای ۲۰۰ نفر سیگاری و غیر سیگاری در حالت تعادل باید ۳۱۵ نفر با **C** برابر ۴ بگیریم تا خطاهای آماری برابر شود. در این صورت بایست به محقق گفته شود که شما نمونه‌گیری را به صورت متوالی باید تا نفر ۳۱۵ام ادامه دهید تا در صورت سیگاری بودن حداقل ۲۰٪ ایشان، به حداقل ۶۳ نفر سیگاری و در مقابل ۲۵۲ نفر غیر سیگاری دست پیدا کنید.

اگر با این فرمول کمی تمرین کنید متوجه می‌شوید که اگر

مقدار c بزرگ شود، کم کم تأثیرش در کاهش حجم نمونه گروه کوچکتر کمتر شده ولی در عوض به دلیل زیاد شدن سریع حجم نمونه گروه بزرگ‌تر، در کل حجم نمونه افزایش قابل ملاحظه‌ای می‌یابد. این واقعیت در نمودار شماره ۱ نمایش داده شده است. در این نمودار نشان داده شده اگر حجم نمونه کلی در حالت تعادل برابر ۵۰۰ نفر باشد، وقتی مقدار C به ۱۰ می‌رسد در کل بایست بیش از ۱۵۰۰ نفر وارد مطالعه شوند تا خطاهای آماری ثابت بمانند



نمودار شماره ۱- تأثیر مقدار C (نسبت بیماران در دو گروه) بر حجم نمونه کل

با افزایش مقدار C حجم نمونه کل به سرعت افزایش می‌یابد در حالی که کاهش حجم نمونه در گروه کوچک‌تر (n1) به صورت تقریباً نامحسوس کاهش می‌یابد.

البته در شرایط دیگری نیز ممکن است مجبور شویم نوع نمونه گیری را به شکلی ترتیب دهیم که از حالت تعادل خارج شود. به عنوان مثال در مطالعات مورد- شاهدی، در بسیاری موارد گرفتن موارد بسیار دشوارتر از گرفتن شاهدها است چراکه معمولاً از تکنیک مورد- شاهدی برای مطالعه بر روی بیماری‌های نادر استفاده می‌شود و در نتیجه گرفتن موارد سخت‌تر می‌باشد. در این گونه مطالعات گاه محقق تصمیم می‌گیرد برای هر مورد مثلاً دو یا حتی بیشتر شاهد بگیرد. برای چنین مواردی به راحتی می‌توان از فرمول فوق استفاده نمود و حجم نمونه را برای مقادیر متفاوت C محاسبه کرد. مثلاً اگر در حالت تعادل بایست ۱۰۰ مورد و ۱۰۰ شاهد وارد مطالعه شوند، اگر بخواهیم به ازای هر مورد، دو شاهد بگیریم یعنی مقدار c را برابر ۲ در نظر بگیریم، بایست ۷۵ مورد و در مقابل ۱۵۰ شاهد وارد مطالعه کنیم. با توجه به این محاسبات محقق باید ببیند ارزش دارد که به ازای کاهش ۲۵ نفری در حجم نمونه گروه مورد، ۵۰ شاهد بیشتر وارد مطالعه کند. اگر زحمت گرفتن هر مورد بیش از دو برابر زحمت گرفتن هر شاهد باشد، قطعاً چنین تعدیلی توصیه می‌شود.

از موارد دیگر کاربرد این فرمول زمانی است که در کارآزمایی‌های بالینی هدف، مقایسه یک داروی جدید با یک داروی استاندارد است و به دلایل اخلاقی و ندانستن شکل اثر داروی جدید و یا عوارض ناخواسته آن، محقق سعی می‌کند تا تعداد نمونه‌هایی که داروی جدید را دریافت می‌کنند را به حداقل برساند و در عوض اشکال زیادی برای گرفتن نمونه بیشتر از گروه دیگر یعنی گروهی که داروی استاندارد را دریافت می‌کنند وجود ندارد. در این موارد نیز به راحتی می‌توان از این فرمول برای تعدیل حجم نمونه استفاده نمود و با کاهش حجم نمونه در گروه مداخله جدید، به حجم نمونه در گروه دریافت کننده داروی استاندارد افزود.

تصحیح حجم نمونه برای مقادیر نامشخص

تقریباً کمتر تحقیقی را می‌توان یافت که تمام موارد گزینش شده تا انتهای مطالعه باقی بمانند و بتوان تمام اطلاعات آن‌ها را وارد مطالعه کرد. دلایل زیادی وجود دارد که تعدادی از نمونه‌ها در هنگام آنالیز در مطالعه وارد نشوند. در مطالعات طولی که نمونه‌ها در طول زمان پیگیری می‌شوند، گم شدن و یا خروج نمونه‌ها به دلیل عدم رضایت به ادامه کار و یا مواجه شدن با شرایط غیر عادی و غیر معمول امری رایج است و هر چه زمان پیگیری طولانی‌تر شود میزان حذف نیز افزایش می‌یابد. چنین پدیده‌ای هم در مطالعات طولی مشاهده‌ای مانند مطالعات هم‌گروهی پیش می‌آید و هم در مطالعات مداخله‌ای که محقق خود افراد را در گروه‌های مختلف تقسیم می‌کند و در طول زمان اثر مداخلات خود را بررسی می‌کند.

البته حذف نمونه‌ها مختص مطالعات طولی نیست و در مطالعات مقطعی نیز دیده می‌شود. عدم پاسخگویی به همه یا تعدادی از سوالات و یا اشکال در انجام آزمایشات از جمله عللی است که باعث بروز موارد حذف شده در مطالعات مقطعی می‌شود. حتی گاه محقق در زمان تجزیه و تحلیل برای افزایش کیفیت اطلاعات تصمیم می‌گیرد بعضی از نمونه‌ها را از مطالعه خارج کند؛ نمونه‌ای از چنین تصمیمی در زمان تجزیه و تحلیل، حذف نمونه‌های خارج از دامنه قابل قبول (out of range) است.

با توجه به توضیحات فوق برای تعدیل حجم نمونه محاسبه شده، بایست نقش و تاثیر مقادیر نامشخص در انتهای مطالعه را از ابتدا در نظر گرفت. به عبارتی باید درصدی بیش از حجم نمونه

محاسبه شده وارد مطالعه نمود تا در صورت حذف بعضی از نمونه‌ها در طول مطالعه در انتها تعداد نمونه‌ها به اندازه کافی باشد. ولی قبل از پرداختن به شیوه تعدیل حجم نمونه، باید توجه ویژه به این نکته داشت که اینگونه تعدیل‌ها فقط زمانی کاربرد دارند که نمونه‌های نامشخص فقط بر اساس شانس و تصادف حذف شده باشد و هیچ‌گونه سوگیری در نمونه‌گیری ایجاد نشده باشد (۴) در صورت حذف غیر تصادفی نمونه‌ها هر میزان حجم نمونه تعدیل گردد نمی‌تواند خطای مربوطه را کاهش دهد و تنها راه ممکن که تا حدودی اثرات سوگیری را ممکن است کاهش دهد انجام bias analysis است که شرح آن خارج از موضوع این مقاله است (۵).

برای روشن‌تر شدن مطلب به مثالی درباره تصحیح حجم نمونه برای مقادیر نامشخص توجه نمایید. فرض کنید که برای تعیین فراوانی مصرف سیگار در میان دانش‌آموزان حجم نمونه محاسبه شده ۶۰۰ نفر بوده است. بعد از مطالعه و وارد نمودن ۶۰۰ دانش‌آموز، پاسخ‌های ۱۰۰ نفر از ایشان غیر قابل قبول بوده و یا به بعضی از سوالات کلیدی پاسخ نداده‌اند. بر اساس پاسخ ۵۰۰ دانش‌آموز باقی مانده، تخمین فراوانی سیگار کشیدن ۱۳٪ بدست آمده است. آیا به نظر شما خطای حاصل از حذف ۱۰۰ نفر فقط کاهش دقت در برآورد این فراوانی و زیاد شدن خطای تصادفی خواهد بود؟ جواب روشن است، خیر؛ نمونه‌های بدون پاسخ احتمالاً بیشتر از میان کسانی بوده‌اند که از بیان رفتار سیگار کشیدن خود طفره رفته‌اند. به عبارتی حذف نمونه‌ها بر اساس تصادف نبوده و فراوانی سیگار کشیدن در نمونه‌های حذف شده با نمونه‌های باقی مانده تفاوت قابل ملاحظه‌ای داشته است. در این صورت حتی اگر از ابتدا به جای ۶۰۰ نمونه، ۶۰۰۰ نمونه نیز گرفته می‌شد، خطای منظم و سوگیری مذکور در نمونه‌های دارای پاسخ کماکان وجود می‌داشت و با افزایش نمونه فقط فراوانی سوگیرانه‌ای را با خطای تصادفی کمتر می‌توان پیش‌بینی نمود. با توجه به این توضیحات، بالا بردن حجم نمونه هرگز نمی‌تواند خطای منظم حاصل از سوگیری در نمونه‌گیری را از بین ببرد و دقت در نمونه‌گیری دقیق و استاندارد از اهم نکاتی است که باید در یک تحقیق به آن توجه خاص مبذول داشت.

حال برای حذف اثرات مقادیر نامشخص تصادفی چه باید کرد؟ در نظر بگیرید که کل حجم نمونه n برای مطالعه نهایی مورد نیاز بوده و یک نسبت (q) از شرکت کردن در مطالعه امتناع ورزیده و یا از مطالعه حذف می‌شوند. در این حالت کل تعداد مواردی که بایستی از ابتدا وارد مطالعه شوند از فرمول زیر بدست می‌آید (۳).

$$\hat{n} = \frac{n}{(1-q)}$$

به عنوان مثال تصور نمایید که ۱۰ ٪ از موارد در یک کار آزمایی درمانی از همان ابتدا از شرکت در مطالعه امتناع ورزند، اگر حجم نمونه اولیه برای این مطالعه۵۲۰ نفر محاسبه شده باشد، برای پوشش خطای از دست دادن نمونه‌ها بایست از فرمول زیر استفاده نمود.

$$\hat{n} = \frac{520}{(1-0.1)} = \frac{520}{0.9} = 578$$

بنابراین حدود ۵۷۸ افراد واجد شرایط در ابتدا باید وارد مطالعه شوند. البته تخمین q یعنی درصد افراد بدون پاسخ در بسیاری از مطالعات در ابتدای مطالعه بسیار دشوار است و لذا سعی می‌گردد به صورت محافظه کارانه این درصد را بالاترین حد مورد انتظار در نظر بگیرند و بدین شکل حجم نمونه را کمی بیش از حد محاسبه کنند تا در انتهای مطالعه با کمترین مشکل مواجه شوند. چنین رویکردی اگرچه احتمالاً کمترین اشکال را در زمان تجزیه و تحلیل اطلاعات به وجود خواهد آورد، ولی همیشه نیز مطلوب نیست به خصوص زمانی که نمونه‌گیری دشوار و به خصوص هزینه بر باشد. در این گونه موارد بایستی سعی گردد با انجام مطالعات پایلوت و یا استفاده از تجربیات دیگران تخمین بهتری از q را بدست آورد.

نکته آخر در خصوص نمونه‌های مقادیر نامشخص، حداکثر قابل قبول برای q است؛ به فرض نبودن سوگیری در نمونه‌گیری، آیا با تصحیح فوق می‌توان برای یک q بالا به نتیجه مطلوب دست یافت؟ جواب منفی است. به عبارتی اگر مقدار q بالا باشد و موارد بدون پاسخ و یا حذف شده از حدی بیشتر باشد، نمی‌توان با تکیه بر این فرمول حجم نمونه را اصلاح کرد. اما حداکثر مقدار قابل قبول برای q چه میزان است؟ در این خصوص اتفاق نظر کلی آن است که q نباید بیش از ۳۰٪ باشد (۳). به عبارتی حذف بیش از ۳۰ درصد نمونه‌ها در زمان تجزیه و تحلیل اطلاعات مشکل آفرین خواهد شد. البته این درصد معمولاً برای مطالعات پرسشنامه‌ای در نظر گرفته می‌شود و برای مطالعات کارآزمایی بالینی معمولاً کار سخت‌گیرانه‌تر بوده و درصد مذکور را کمتر در نظر می‌گیرند. به همین دلیل و در زمان‌هایی که واقعاً راهکاری برای کاهش موارد نامشخص وجود ندارد باید برای تحلیل دقیق‌تر نتایج با استفاده از روش‌های خاص موارد بدون پاسخ و یا نامشخص را ابتدا برآورد نمود و سپس تجزیه و تحلیل‌های نهایی را انجام داد تا خطای مذکور را به حداقل ممکن رساند.

برآورد حجم نمونه برای روش نمونه‌گیری چند مرحله‌ای

(multistage sampling) در بسیاری مواقع به منظور کاهش هزینه‌های مطالعه و سایر ملاحظات اجرایی محققین از روش نمونه گیری چند مرحله‌ای استفاده می‌کنند. این در حالی است که فرمول‌های بیان شده تا بدین جا با فرض نمونه‌گیری تصادفی ساده می‌باشند.

کارآیی نمونه‌گیری‌ها با توجه به نمونه‌گیری تصادفی ساده محاسبه می‌شود و بر اساس یک اصل کلی دقت و کارایی نمونه‌گیری طبقه‌ای (stratified) بهتر و نمونه‌گیری خوشه‌ای و چند مرحله‌ای کمتر از نمونه‌گیری تصادفی ساده است. برای نشان دادن این میزان کارآیی از شاخصی بنام اثرطرح (۴) استفاده می‌کنند و برای تعدیل حجم نمونه، ابتدا بر اساس فرمول‌های مربوط به نمونه‌گیری تصادفی ساده حجم نمونه را حساب و بعد در اثر طرح ضرب می‌کنند. به عنوان مثال اگر در یک مطالعه چند مرحله‌ای اثرطرح برابر ۱/۶ باشد و مثلاً فرمول برآورد یک نسبت حجم نمونه را ۸۰۰ نفر تعیین نموده باشد، برای تعدیل حجم نمونه باید مقدار حجم نمونه را ۶۰٪ افزایش داد (۸۰۰*۱.۶=۱۲۸۰).

اما سوالی که بایست به آن جواب داد مفهوم مقدار اثرطرح و شیوه سنجش آن است. در واقع این شاخص نشان می‌دهد که میزان همبستگی درون گروهی نسبت به همبستگی بین گروهی چه میزان است. برای روشن‌تر شدن این مفهوم بهتر است از یک مثال استفاده شود. فرض کنید می‌خواهیم میزان شیوع دیابت را در شهر کرمان بسنجیم و برای این کار ۱۰۰۰ نفر از ۵۰ خوشه به صورت تصادفی از شهر انتخاب و در هر خوشه ۲۰ نفر از نظر دیابت مورد بررسی قرارگرفته‌اند. در این مطالعه هر خوشه متشکل از چندین خانواده در نظر گرفته می‌شوند که مجاور هم زندگی می‌کنند و تمام افراد بالای ۲۰ سال آن‌ها وارد مطالعه شده‌اند. حال این سوال پیش می‌آید که افراد درون هر خوشه چه میزان در مورد متغیر مورد بررسی یعنی دیابتی بودن با هم تشابه دارند؟ در حالت فرضی ممکن است این همبستگی ۱۰۰٪ باشد، به این معنی است که اگر نفر اول یک خوشه را بررسی کنیم با دقت ۱۰۰٪ می توانیم بگوییم که سایر ۱۹ نفر آن خوشه دیابتی هستند یا خیر؟ به عبارتی اگر نفر اول دیابتی باشد حتماً و حتماً ۱۹ نفر دیگر نیز دیابتی خواهند بود و بالعکس اگر نفر اول دیابتی نباشد، سایر افراد نیز دیابتی نخواهند بود. اگرچه در عمل چنین همبستگی ۱۰۰٪ تقریباً بسیار بعید است، ولی به صورت تئوریک آیا واقعاً در این مثال هر نمونه به اندازه یک نمونه مستقل اطلاعات وارد مطالعه می‌کند؟ پرواضح است که جواب منفی است و بهتر است بگوییم ۱۰۰۰ نفر با این شیوه نمونه‌گیری هیچ حرف بیشتری از گرفتن ۵۰ نمونه تصادفی از شهر نخواهد زد چراکه همبستگی درون گروهی ۱۰۰٪ است و افراد درون خوشه اطلاعات اضافه‌ای وارد مطالعه نمی‌کنند.

در مقابل اگر نمونه‌های درون خوشه هیچ تناسبی با هم نداشته باشند و شانس و احتمال دیابتی بودن دو نفر در یک خوشه برابر احتمال دیابتی بودن دو نفر مستقل و تصادفی از شهر کرمان باشد، در این صورت همبستگی درون گروهی صفر است و عملاً هر نمونه مانند یک نمونه مستقل تصادفی از جامعه عمل نموده و در این حالت اثرطرح برابر یک است. به عبارتی ۱۰۰۰ نمونه تصادفی ساده از این شهر همان قدر اطلاعات وارد مطالعه می‌کنند که ۵۰ خوشه ۲۰ نفری اطلاعات می‌دهند.

با توجه به توضیحات فوق می‌توان باور نمود که در عمل احتمالاً میزان اثرطرح بیش از یک است چراکه افرادی که در یک خوشه قرار دارند از نظر ژنتیک (در یک خانواده) و از نظر اقتصادی- اجتماعی و شیوه زندگی (در یک خانواده و در همسایه‌ها) با یکدیگر تشابه دارند و به همین دلیل شانس دیابتی بودن در افراد یک خوشه بیش از افراد مستقل در یک جامعه است.

با توجه به این توضیحات هر قدر مشابهت افراد یک خوشه به یکدیگر نزدیک شود، مقدار اثرطرح افزایش یافته و با تعدیل حجم نمونه، باید افراد بیشتری وارد مطالعه نمود. این میزان تشابه را با ρ (با تلفظ rho) نشان می‌دهند که نشان دهنده همبستگی درون خوشه‌ای (intra-cluster correlation) است و برای محاسبه آن بایست از نتایج مطالعات مشابه و یا مطالعه پایلوت کمک گرفت. در صورتی که داده هایی از یک مطالعه مشابه یا مطالعه مقدماتی موجود باشد می‌توان با استفاده از تکنیک آنالیز واریانس و فرمول زیر همبستگی مشاهدات در هر خوشه را تخمین زد (۵).

$$\rho = (MSB-MSW)/ \{1+ (m-1)\ MSW\}$$

MSB= between cluster mean square

MSW= within cluster mean square

البته مقدار اثرطرح علاوه بر مقدار ρ به اندازه نمونه‌های هر خوشه نیز وابسته است. به عبارتی اگر تعداد نمونه‌های یک خوشه زیاد شود، با فرض ثابت بودن ρ، مقدار اثرطرح نیز افزایش می‌یابد. فرمول مورد استفاده برای نشان دادن این ارتباط به شرح زیر است (۶).

$$\text{Design effect} = 1+ (m-1)*\rho$$

در فرمول فوق m نشان دهنده تعداد افراد در هر خوشه می‌باشد. در اینجا فرض شده است که تعداد افراد در خوشه‌های مختلف برابر است. با توجه به فرمول فوق اگر مقدار ρ برابر ۰/۰۴ باشد، مقدار اثر اگر خوشه‌ها

۱۰ نفره انتخاب شوند برابر ۱/۳۶ خواهد بود و به عبارتی بایست حجم نمونه محاسبه شده با فرمول‌های ساده را ۳۶٪ افزایش داد، ولی اگر اندازه خوشه‌ها به ۳۰ نفر برسد مقدار اثر طرح ۲.۱۶ خواهد شد که نشان می‌دهد حجم نمونه باید ضربدر ۲/۱۶ شود (۱۱۶٪ افزایش) و لذا میزان حجم نمونه در چنین نمونه‌گیری خوشه‌ای بسیار افزایش خواهد یافت. این نکته مبین این حقیقت است که در نمونه‌گیری‌های خوشه‌ای نباید اندازه خوشه‌ها را بزرگ نمود چرا که میزان اثر و در نتیجه حجم نمونه مورد نیاز به شدت افزایش خواهد یافت.

به عنوان یک راهکار کلی می‌توان بیان نمود که معمولاً مقدار اثر طرح در نمونه‌گیری طبقه‌ای بین ۰/۸ تا ۱ و در نمونه‌گیری خوشه‌ای ساده حداقل ۱/۵ است و در نمونه‌گیری چند مرحله‌ای اگر اندازه خوشه‌های بزرگ دیده نشود معمولاً مقدار اثرطرح بین ۱/۳ تا ۲ خواهد بود (۲،۵). در کل می توان بیان نمود که فرمول حجم نمونه مؤثر Effective Sample Size (ESS)) برای نمونه‌گیری خوشه‌ای را می‌توان به طریق زیر به دست آورد.

$$ESS = \frac{mk}{DE}$$

به گونه‌ای که m برابر با تعداد افراد در هر خوشه بوده و k تعداد خوشه‌ها است. فرمول DE یا همان اثر طرح در بالا شرح داده شده است.

برای بیان اثر ρ و Design effect بر روی حجم نمونه و power در جداول ۳،۲ و۴ مثالی ارایه شده است. فرض کنید که در یک طرح تحقیقاتی می خواهیم به بررسی تعدادی از بیماران بپردازیم واز چندین پزشک درخواست کرده‌اید تا تعدادی از بیمارانشان را معرفی کنند. در اینجا هر یک از پزشکان به عنوان یک خوشه و تعداد بیماران موجود در هر خوشه نمونه‌های ما را تشکیل می‌دهند. فرض نماییم که در اینجا اثر ρ برابر با با ۰/۰۱۷ باشد.

جدول شماره ۲ نشان دهنده تغییرات در اندازه نمونه مؤثر و قدرت است، در حالی که ما برای mk ثابت تعداد بیماران و پزشکان متفاوتی داشته باشیم. جدول شماره ۳ نشان دهنده اثر افزایش m بوده در حالیکه k ثابت نگه داشته شده باشد. به افزایش در اثر طرح با افزایش m و تاثیر آن بر روی حجم نمونه مؤثر توجه نمایید.

در این حالت پژوهشگر بایستی ۸۰ بیمار را به ازای هر پزشک

جدول شماره ۲- حجم نمونه موثر و قدرت آن در صورتیکه کل افراد و تعداد بیماران ثابت باشند.

Power	ρ= ۰/۰۱۷		کل افراد (mk)	تعداد بیماران (m)	تعداد پزشکان (k)
	ESS	DE			
۶۱	۸۴	۱/۵۲۷	۱۲۸	۳۲	۴
۷۰	۱۰۲	۱/۲۵۵	۱۲۸	۱۶	۸
۷۵	۱۱۴	۱/۱۱۹	۱۲۸	۸	۱۶
۷۸	۱۲۲	۱/۰۵۱	۱۲۸	۴	۳۲
۷۹	۱۲۶	۱/۰۱۷	۱۲۸	۲	۶۴
۸۰	۱۲۸	۱/۰۰۰	۱۲۸	۱	۱۲۸

جدول شماره ۳- حجم نمونه موثر و قدرت آن در صورتیکه تعداد پزشکان ثابت باشند.

Power	ρ= ۰/۰۱۷		کل افراد (mk)	تعداد بیماران (m)	تعداد پزشکان (k)
	ESS	DE			
۲۹	۳۴	۱/۱۵۳	۴۰	۱۰	۴
۴۷	۶۰	۱/۳۲۳	۸۰	۲۰	۴
۶۷	۹۶	۱/۶۶۳	۱۶۰	۴۰	۴
۸۲	۱۳۶	۲/۳۴۳	۳۲۰	۸۰	۴

جدول شماره ٤- حجم نمونه مؤثر و قدرت آن در صورتیکه تعداد بیماران ثابت باشند.

Power	ρ= ۰/۰۱۷		کل افراد (mk)	تعداد بیماران (m)	تعداد پزشکان (k)
	ESS	DE			
۱۶	۱۸	۱/۱۵۳	۲۰	۱۰	۲
۳۰	۳۶	۱/۱۵۳	۴۰	۱۰	۴
۵۰	۷۰	۱/۱۵۳	۸۰	۱۰	۸
۸۳	۱۳۸	۱/۱۵۳	۱۶۰	۱۰	۱۶

(در مجموع ۳۲۰ مورد) به مطالعه وارد کرده تا به قدرت مناسب با ۴ پزشک برسد.

در جدول شماره ۴ نشان داده شده است که با افزایش تعداد پزشکان به ۱۶ تنها نیاز به ۱۶۰ بیمار بوده تا به قدرت بالای ۸۰ ٪ برسیم (۷).

برآورد حجم نمونه برای آزمون‌های ناپارامتری

از موارد دیگری که در محاسبه حجم نمونه بایست مد نظر قرار گیرد، نوع آزمون آماری است که برای تحلیل نهایی داده‌ها مورد استفاده قرار می‌گیرند. به صورت بسیار ساده، می‌توان آزمون‌ها را به پارامتریک و غیرپارامتریک تقسیم نمود. در آزمون‌های پارامتریک مانند t-test، ANOVA، Pearson correlation و linear regression، پیش فرض‌هایی در مورد، نوع متغیرهای عددی (یعنی فاصله‌ای و یا نسبتی بودن)، توزیع و واریانس آن‌ها در نظر گرفته می‌شود و محاسبات با این پیش فرض‌ها صورت می‌گیرد. در زمانی که این پیش فرض‌ها محقق نباشند و به راحتی

و با تبدیل داده‌ها (data manipulation such as transformation) نتوانیم این پیش فرض‌ها را محقق نماییم، بایست از روش‌های آماری جایگزین که اصلاحاً روش‌های غیرپارامتریک نام دارند استفاده نمود.

محاسبات حجم نمونه با فرمول‌های بیان شده برای مقایسه میانگین‌ها، برای زمانی است که آزمون‌های نهایی با استفاده از روش‌های پارامتریک صورت می‌گیرند، لذا اگر مجبور باشیم از روش‌های غیر پارامتریک استفاده کنیم، بایست حجم نمونه‌ها را تعدیل کنیم. لازم به ذکر است که چنین تعدیلی برای محاسبه حجم نمونه برای مقایسه دو نسبت لازم نیست چرا که از نظر ماهیت مقایسه دو نسبت با استفاده از روش‌های غیرپارامتریک صورت

می‌گیرد و امکان استفاده از آزمون‌های پارامتریک به راحتی ممکن نیست.

از آنجایی که قدرت و توان آزمون‌های آماری غیرپارامتریک کمتر از آزمون‌های آماری پارامتریک است، لذا برای رسیدن به دقت کافی باید حجم نمونه افزایش یابد. به عبارتی اگر با دقت آماری مشخص و از قبل تعیین شده برای آزمون آماری پارامتریک باید در هر گروه ۱۰۰ نمونه وارد مطالعه شود، زمانی که شما مجبور به استفاده از آزمون غیرپارامتریک می‌شوید، باید حجم نمونه بیشتری را بگیرید تا در نهایت تجزیه و تحلیل شما با همان دقت قبلی باشد.

حال این سوال پیش می‌آید که جهت تصحیح برای آزمون‌های غیرپارامتریک چه میزان حجم نمونه بایست افزایش یابد؟ پاسخ این سوال بر می‌گردد به اینکه که توان آزمون غیرپارامتریک چه میزان کمتر از توان آزمون پارامتریک معادل است. برای نشان دادن این موضوع شاخصی تعریف می‌شود به نام کارآمدی توان آزمون غیرپارامتریک (۸-۵). مثلاً اگر این توان برای یک آزمون غیرپارامتریک ۹۵٪ باشد، یعنی باید حجم نمونه محاسبه شده در معکوس ۹۵٪ ضرب شود. به عبارتی باید حجم نمونه محاسبه شده تقریبا ۵.۲٪ افزایش یابد تا در انتها خطاهای آماری در تحلیل ثابت باقی بمانند.

در جدول ذیل (جدول شماره ۳) کارآمدی توان بعضی از مهم‌ترین آزمون‌های آماری غیرپارامتریک در مقابل معادل پارامتریک آن‌ها بیان شده است.

جدول شماره ۳- کارآمدی توان انواع مهم آزمون‌های غیرپارامتریک در مقابل معادل پارامتریک خود

آزمون غیرپارامتریک	معادل پارامتریک	توان کارآمدی (٪)
Mann-Whitney	Pooled t-test	95.5
Wilcoxon Signed-rank test	paired t-test	95.5
Wilcoxon One sample test	One sample t test	95.5
Kruskal-Wallis	ANOVA	95.5
Spearman or Kendal's correlation	Pearson correlation	91
Kendal correlation	Pearson correlation	91
Freedman	Repeated ANOVA	$\frac{0.995k}{k+\square}$ ▲

▲ در اینجا k عبارت است از تعداد گروه‌ها

مورد تأمل آن است که گاهی محققین متغیرهای کمی را به حالت کیفی دو حالته تبدیل می‌کنند. این امر روی روش تحلیل داده‌ها تأثیر می‌گذارد. به عنوان مثال در جایی که امکان استفاده از Paired t-test برای تحلیل متغیر عدی وجود دارد گروه‌بندی کردن متغیر، محقق را ناچار می‌کند که از آزمون Mc-Nemar استفاده کند. این امر روی کارآمدی توان مطالعه تأثیر مستقیم دارد و باعث کاهش کارآمدی آزمون آماری می‌شود و در شرایط خاص ممکن است کارآمدی توان آزمون حتی به ۶۲٪ نیز برسد به همین دلیل در تجزیه و تحلیل آماری چنین تقسیم‌بندی به صورت عمومی توصیه نمی‌شود (۸,۹). بحث در مورد مزایا و معایب گروه بندی خارج از حوصله این مقاله است.

برآورد حجم نمونه برای جامعه محدود

درصورتیکه جامعه مورد تحقیق ما یک جامعه محدود یا جامعه‌ای بدون جایگزینی باشد پس از به دست آوردن حجم نمونه بایستی آن را با استفاده از فرمول زیر تصحیح نمود (۱۰).

$$n = \frac{n_0\ N}{n_0\ + (N-1)}$$

فرض کنید که حجم نمونه به دست آمده برای یک تحقیق ۲۰۰ باشد، در صورتیکه کل جامعه مورد تحقیق ما ۵۰۰۰ =N باشد، در این صورت حجم نمونه نهایی مورد استفاده ما عبارت است از:

$$n = \frac{(200) * (5000)}{(200) + (5000 - 1)} = 192.3$$

در اینجا می‌توان مشاهده نمود، در صورتیکه اندازه جمعیت ما در مقایسه با کل جمعیت هدف کوچک باشد، استفاده از این فرمول حجم نمونه را اندکی کاهش می‌دهد که این امر بدین خاطر

است که یک اندازه نمونه مشخص به نسبت اطلاعات بیشتری را برای جمعیت‌های کوچک تر نسبت به جمعیت بزرگتر فراهم می‌کند. البته بایستی به این نکته نیز توجه نمود که اگر نسبت بین نمونه به کل جمعیت بیش از ۵٪ بود نیاز به تعدیل وجود دارد و در غیر این صورت تعدیل، حجم نمونه را تغییر محسوسی نمی‌دهد.

نتیجه‌گیری

در این مقاله سعی بر آن شد که برای تعیین حجم نمونه در شرایط خاص مباحثی ارایه شود و تا حدودی نکات مربوط به نمونه‌گیری با تکنیک‌های ساده و قابل استفاده ارایه شود. ولی شرایط بسیار خاص دیگری نیز وجود دارد که در گنجایش این مطلب نبوده است. از جمله این نکات می‌توان به بحث حجم نمونه در مطالعات کیفی و مفهوم اشباع شدن، تأثیر جور سازی بر حجم نمونه و روش‌های تعیین نمونه در مدل‌های رگرسیونی اشاره کرد. نکته دیگر آن است که بعضی اوقات از تعداد محدودی نمونه در دسترس به جای نمونه گیری استفاده می‌شود. در اینگونه موارد لازم است بجای محاسبه حجم نمونه به محاسبه توان مطالعه پرداخت تا بتوان کیفیت مطالعه را مورد ارزیابی قرار داد. به یاری خدا این نکات در یک مقاله مستقل مورد بحث قرار خواهند گرفت. امید است نکات ارایه شده در این مقاله محققین را با بعضی از مباحث در روش‌های تعیین نمونه بیش از پیش آشنا کرده باشد.

منابع

1. Haghdoost, A.A., Do You Want to Gain a Profound Insight into Sample Size and Statistical Power. Iranian journal of epidemiology, 2008; 5: 57-63.
2. Machin, D., Sample size tables for clinical studies. 1997; Wiley-Blackwell.
3. Whitley E. and Ball J. Statistics review 4: sample size calculations. Critical Care, 2002; 6: 335.
4. Katz J and Zeger SL. Estimation of design effects in cluster surveys. Annals of Epidemiology, 1994; 4: 295-301.
5. Woodward, M., Epidemiology: study design and data analysis. 1999; CRC Press : 420.92.
6. Prajapati B, Dunne M, and Armstrong R. Sample size estimation and statistical power analyses. Optometry Today. 2010(July).
7. Killip S, Mahfoud Z, and Pearce K. What is an intracluster correlation coefficient? Crucial concepts for primary care researchers. The Annals of Family Medicine, 2004; 2: 204.
8. Siegel, S., Nonparametric statistics. American Statistician, 1957; 11: 13-19.
9. Nasirian M, Sadeghi M, and Haghdoost AA. Principal for information analysis and correctly issued result. Isfahan university of medical science, 2009; 27: 646-59.
10. Amidi A and Meshkati MR. Theory of sampling and application of it. Center publication of university, 2007; 1:92.

Iranian Journal of Epidemiology 2011; 7(2): 67-74.

Review Article

How to Estimate the Sample Size in Special Conditions? (Part two)

Haghdoost A[1], Baneshi MR [2], Marzban M[3]

1- PhD of Epidemiology Physiology research center, Kerman University of Medical Sciences, Kerman, Iran
2- PhD of Biostatistics. Department of Biostatistics and Epidemiology, Kerman University of Medical Sciences, Kerman, Iran
3- Department of Biostatistics and Epidemiology, Kerman University of Medical Sciences, Kerman, Iran
Corresponding author: Marzban M., marzbanh@gmail.com

In the previous paper, the basic concepts of sample size calculation were presented. This paper explores main post-calculation adjustments of the sample size calculation in special circumstances such as multiple group comparisons, unbalanced studies (with unequal number of subjects in different groups); sample size correction for missing data, and adjustment for finite population size. In addition, the concept of design effect in multi-stage sampling and its impact on the sample size are presented. We then focused on the sample size estimation when we have to use non-parametric statistical tests for data analysis. The concept of power-efficacy of parametric versus non-parametric methods and its use in the correction of sample size has been explained.

Keywords: Sample size, Multiple comparison, Design effect, Non-parametric methods, Missing data, Finite population size