

آیا می‌خواهید مفهوم حجم نمونه و توان آزمون‌های آماری را عمیق‌تر درک نمایید؟

علی اکبر حقدوست^۱

^۱ دانشیار، مرکز تحقیقات فیزیولوژی، دانشگاه علوم پزشکی کرمان، ایران

نویسنده رابط: علی اکبر حقدوست، نشانی: کرمان، دانشگاه علوم پزشکی کرمان مرکز تحقیقات فیزیولوژی. تلفن: ۰۲۱۱۳۰۰۵-۰۳۴۱، پست الکترونیک: ahaghdooost@gmail.com

تاریخ دریافت: ۱۳۸۷/۰۵/۲۰ پذیرش: ۱۳۸۷/۱۰/۰۷

مبحث حجم نمونه و نحوه محاسبه آن یکی از مباحث بسیار کلیدی در تحقیقات علوم پزشکی است که معمولاً محققین و حتی خوانندگان مقالات در این خصوص سوالات بسیاری دارند و این موضوع نشان می‌دهد که باید مفاهیم در این خصوص به زبان ساده ولی دقیق شرح داده شوند.

در این مقاله سعی خواهد شد با زبانی بسیار ساده و با کمترین پیچیدگی آماری، این مبحث برای خوانندگان غیرآشنا با مفاهیم پیچیده آماری بیان و با ارایه مثال‌های کاملاً مفهومی، عوامل موثر بر حجم نمونه شرح داده شوند. در این مقاله ابتدا در مورد مفاهیم پایه در محاسبه حجم نمونه مانند مفهوم میزان اثر (effect size)، مفهوم دامنه و ضریب اطمینان و خطاهای آماری بحث شده و مفروضات مورد نیاز برای محاسبه حجم نمونه شرح داده خواهند شد. سپس به بیان فرمول محاسبه حجم نمونه برای برآورد یک میانگین و یا یک نسبت و همچنین فرمول رایج برای محاسبه حجم نمونه در مقایسه دو میانگین و یا دو نسبت پرداخته خواهد شد. در ادامه این سری از مقالات نیز سعی خواهد شد به زوایای فنی‌تر موضوع پرداخته شده و راهکارهای ساده شده‌ای برای محاسبه حجم نمونه در شرایط خاص و تصحیحات مربوطه ارایه گردد.

کلمات کلیدی: حجم نمونه، متدولوژی تحقیق، محاسبات آماری، بررسی نقادانه

شناخت عمیق‌تر مفهوم حجم نمونه و عوامل موثر بر آن

برای پاسخ به سوالات پژوهش باید سعی گردد تا متغیرهای اصلی و مورد نظر در گروه مناسبی از جمعیت هدف مورد سنجش قرار گیرد. این گروه باید نمونه خوب و مبین (representative) باشد و به عبارتی سوگیری درنمونه‌گیری (selection bias) وجود نداشته باشد. البته نکته دوم آن است که اندازه و حجم نمونه گرفته شده باید کافی باشد. تعداد کم و زیاد حجم نمونه یعنی به هدر دادن بودجه و منابع. مطالعه کوچک و حجم نمونه کم باعث خواهد شد تا دقت نهایی نتیجه‌گیری کم باشد و به عبارتی اعتماد و اعتبار نتایج زیر سوال رود؛ لذا هزینه پرداخته شده و امکانات صرف گردیده است ولی نتیجه‌گیری موثر و ارزشمند بدست نخواهد آمد. در مقابل حجم نمونه بالا نیز باعث خرج شدن زیاد بودجه می‌شود؛ ولی افزایش دقت به اندازه هزینه اضافه شده نخواهد بود ۱-۳.

اما قبل از اینکه در خصوص محاسبه حجم نمونه و عوامل موثر بر آن به صورت کلیشه‌ای مطالبی بیان شود، بد نیست با چند مثال غیرمستقیم در خصوص مفهوم حجم نمونه با هم صحبت کنیم. بگذارید برای آسان شدن مطلب، حجم نمونه را مشابه ذره‌بین بنامیم. اما وجه تشابه ذره‌بین و حجم نمونه در چیست؟ ذره‌بین را بدست می‌گیریم تا بتوانیم اجسام ریز را ببینیم و شناخت بهتری از پیرامون خود داشته باشیم. هرچه ذره‌بین ما قوی‌تر باشد با دقت و قدرت بیشتری می‌توانیم پیرامون خود را بشناسیم. حجم نمونه نیز به ما کمک می‌کند تا در یک تحقیق با قدرت و توان بیشتری بتوانیم ارتباط بین متغیرها را بشناسیم و یا در برآورد شاخص‌ها دقت بیشتری داشته باشیم. اگر قدرت ذره‌بین کم باشد و یا حجم نمونه کوچک در نظر گرفته شود، مسلماً نخواهیم توانست واقعیت‌ها را خوب ببینیم. در مقابل اگر ذره‌بینی بسیار قوی داشته باشیم و یا حجم نمونه را بسیار بالا انتخاب

دارد. مفهوم این عبارت این است که حداقل میزان تاثیر چقدر باید باشد تا از نظر بالینی اهمیت یابد. تصور کنید که واقعاً مصرف رب گوجه‌فرنگی میزان ضریب هوشی را یک دهم درصد افزایش دهد! آیا شما توصیه می‌کنید که فرزندانان زیاد رب گوجه‌فرنگی بخورند تا باهوشتر شوند؟ مسلماً خواهید گفت درست است که واقعاً رب گوجه‌فرنگی هوش را افزایش می‌دهد ولی مقدار آن بسیار کم و تاثیر آن قابل توجه نیست. شاید اضافه نمایید که من زمانی رب گوجه‌فرنگی بیشتری به فرزندانم خواهم داد که مطمئن شوم این تغییر رفتار تغذیه‌ای، حداقل ۲ واحد ضریب هوشی را افزایش می‌دهد. یعنی حداقل میزان تاثیر باید از یک حد معقول بیشتر باشد. این سطح را اصطلاحاً "تفاوت مهم از دیدگاه بالینی" می‌نامیم. یک مثال دیگر، اگر مصرف یک دارو مقدار قند خون را واقعاً به میزان ۰/۱ میلی‌گرم کاهش دهد قطعاً هیچ پزشکی آن را برای درمان دیابت بکار نخواهد برد. بر این اساس، از ابتدا محققین و پزشکان تصور می‌کنند که باید میزان تاثیر دارو حداقل ۲۰ میلی‌گرم در کاهش قند خون باشد تا بتوان به عنوان یک دارو موثر معرفی گردد. این مفهوم حداقل میزان اثر مهم از دیدگاه بالینی است.

بر اساس توضیحات فوق می‌توان بیان نمود که جنس و ماهیت "تفاوت مهم از دیدگاه بالینی" و "میزان اثر" یکی است که اولی را محققین از ابتدا تعیین می‌کنند ولی دومی یک موضوع مجرد از نوع تحقیق است. شدت رابطه بین سیگار و سرطان ریه یک مقدار نسبتاً ثابت است و محققین سعی می‌نمایند در تحقیقات خود میزان آن را تخمین بزنند و بسته به میزان ذاتی این ارتباط، برای کشف آن باید بزرگی مطالعه را تعیین نمود؛ ولی مسلماً اگر این شدت از یک حد مشخصی کوچکتر باشد، شاید از نظر عملی دیگر اهمیت نداشته باشد تا برای کشف آن مطالعه را بسیار بزرگ نماییم چرا که حتی در صورت وجود ارتباط بسیار ضعیف، در عمل چندان مهم نخواهد بود. این سطح را اصطلاحاً حداقل ارتباط و یا تفاوت مهم می‌دانیم.

به مثال اول یعنی شنیدن صدای موسیقی برگردیم. تا اینجا گفته شد که شدت اثر، مشابه شدت صدای موسیقی است. حال برویم سراغ دومین عامل موثر در شنیدن صدای موسیقی، یعنی شدت صداهای مزاحم و پارازیت‌ها. مسلماً هر چه شدت این صداهای مزاحم بیشتر باشد، امکان شنیدن صدای موزون کمتر می‌شود. خوب فکر می‌کنید که شدت صداهای مزاحم در بحث حجم نمونه چه معنایی دارد؟

پاسخ سوال بالا بسیار ساده است. صداهای مزاحم یعنی میزان

نماییم، واقعیت‌های بسیار ریز و کم اهمیت را نیز خواهیم دید و به دلیل دیدن ارتباطات بسیار زیاد ولی کم اهمیت، نمی‌توانیم به دقت در خصوص مهم‌ترین حقایق صحبت کنیم.

بگذارید حجم نمونه را به گونه دیگری نیز توصیف کنیم. تصور نمایید که در یک محیط پر سر و صدا، قصد شنیدن موسیقی دارید؛ آیا در چنین شرایطی می‌توان به راحتی صدای موزون و آهنگ موسیقی را شنید؟ مسلماً کار آسانی نیست. برای درک بهتر صدای موسیقی باید حداقل یکی از سه راهکار زیر را برگزید:

۱- بلند نمودن صدای موسیقی

۲- کاهش صداهای مزاحم

۳- افزایش قدرت شنوایی

حال بگذارید ارتباط این سه روش را با عوامل موثر در تعیین حجم نمونه بررسی کنیم. ابتدا بلند نمودن صدای موسیقی؛ یعنی قوی‌تر شدن شدت عامل اثرگذار. این مفهوم در حجم نمونه تحت عنوان میزان اثر (effect size) و یا در مطالعات پزشکی تحت عنوان "تفاوت مهم از دیدگاه بالینی" (clinical importance difference) شناخته می‌شود.

تصور نمایید ۷۰ سال قبل زندگی می‌کردید و باور داشتید که سیگار یک خطر بزرگ برای انسان‌ها است؛ زمانی که همه مردم و بزرگان برای تفنن سیگار می‌کشیدند و فکر می‌کردند که سیگار کشیدن نه تنها مضر نیست بلکه منافع زیادی هم دارد. حال در این شرایط قصد دارید رابطه بین سیگار کشیدن و بروز انواع سرطان‌ها در بدن را مشخص نمایید. فکر می‌کنید که آیا رابطه بین سیگار و سرطان ریه را راحت‌تر بتوانید کشف نمایید و یا رابطه بین سیگار و سرطان پانکراس را؟ شما بهتر می‌دانید که سیگار عامل خطر قوی‌تری برای سرطان ریه است و لذا احتمالاً با تعداد کمتری نمونه می‌توانید نشان دهید که سرطان ریه در افراد سیگاری بیشتر از غیرسیگاری است. ولی چون شدت ارتباط بین سیگار کشیدن و سرطان پانکراس ضعیف‌تر است، باید جمعیت بزرگتری وارد مطالعه شوند و مدت طولانی‌تری پیگیری شوند تا تفاوت ارزشمندی بین تعداد مبتلایان به سرطان پانکراس در گروه سیگاری و غیرسیگاری کشف شود.

بر اساس توضیحات فوق می‌توان بیان نمود که میزان اثر سیگار بر سرطان ریه بیشتر و قوی‌تر است؛ یعنی صدای موسیقی مربوط به این موضوع بیشتر است و لذا راحت‌تر شنیده و کشف می‌شود. در مقابل چون تاثیر سیگار بر سرطان پانکراس ضعیف‌تر است لذا دیرتر شنیده می‌شود.

اما عبارت "تفاوت مهم از دیدگاه بالینی" مفهوم نسبتاً متفاوتی

بیشتری دارد. شما اگر ۳۰۰۰ نوزاد را وارد مطالعه نمایید احتمالاً تعداد نوزاد مبتلا به کم‌کاری مادرزادی تیروئید کمتر از تعداد انگشتان یک دست خواهد بود. پس بر خلاف نتیجه‌گیری قبلی به نظر می‌رسد که باید برای مطالعه بر روی متغیرهای نادر حجم نمونه بیشتری داشت. نظر شما در خصوص این تناقض چیست؟ می‌دانم که پاسخ به این سوال آسان نیست؛ ولی اگر بتوانید این موضوع را با مفهوم میزان اثر ترکیب نمایید قطعاً به پاسخ مناسب خواهید رسید. البته احتمالاً با خواندن این مقاله تا به انتها به پاسخ خواهید رسید؛ ولی توصیه می‌کنم که چند دقیقه فکر نمایید و سعی کنید پاسخ قانع‌کننده‌ای برای خود بیابید.

خوب تا اینجا در خصوص شدت صدای موسیقی و شدت صداهای مزاحم صحبت کردم. آیا به یاد می‌آورید که سومین عامل در شنیدن صدای موسیقی چه بود؟ بله بالا بردن قدرت شنوایی؛ مسلماً افرادی که گوش تیزتری دارند بهتر و سریعتر می‌توانند صداهای موزون را بشنوند. اما این مفهوم در بحث حجم نمونه چیست؟ پاسخ آسان است. بله پاسخ خود حجم نمونه است. مثال ذره‌بین را بیاد بیاورید؟ بیان کردم که حجم نمونه مشابه ذره‌بین است؛ حال در اینجا می‌گوییم که قدرت شنوایی مشابه حجم نمونه است. حجم نمونه بالا کمک می‌کند که میزان اثر کوچک‌تر نیز کشف شود همانگونه که مثلاً سمعک کمک می‌کند تا اصوات ضعیف‌تر شنیده شود.

پس به صورت خلاصه، شنیدن یک موسیقی (یعنی کشف ارتباط بین متغیرها) به سه عامل وابسته است که عبارت هستند از شدت صدا (میزان اثر)، میزان صداهای مزاحم در محیط (تغییرات تصادفی متغیرها یعنی انحراف معیار) و قدرت شنوایی (حجم نمونه).

با توضیحات داده شده به نظر می‌رسد که زمینه کافی برای ورود به بحث به صورت تخصصی‌تر ایجاد شده است؛ لذا در ادامه به صورت مجزا روش و فرمول‌های محاسبه حجم نمونه شرح داده خواهد شد. لازم به ذکر است که در همه بحث‌های زیر فرض بر این است که جامعه هدف مورد بررسی بزرگ بوده و به زبان آماری بی‌نهایت است. این فرض در بسیاری از تحقیقات وجود دارد و در کل بیان می‌شود که اگر حجم نمونه کمتر از ۵٪ کل جامعه باشد می‌توان فرض بزرگ بودن جامعه آماری را قبول کرد. ولی اگر به هر شکل، فرض مذکور محقق نباشد، باید حجم نمونه محاسبه شده توسط این فرمول‌ها را تعدیل نمود که نتیجه اثر این تعدیل‌ها کم شدن تعداد نمونه‌های مورد نیاز است؛ لذا حتی اگر این فرض مورد توجه واقع نشود، در عمل حجم نمونه محاسبه شده کمی

خطاهای تصادفی و پراکندگی طبیعی متغیرهای مهم مورد سنجش در جامعه و برای متغیرهای عددی این پراکندگی با شاخص انحراف معیار سنجیده می‌شود. اگر به طور طبیعی میزان تغییرات فشارخون در جامعه بسیار متغیر و متفاوت باشد، برای آنکه باور کنیم که یک دارو فشارخون را کاهش می‌دهد، باید حجم نمونه بسیار زیادی را در نظر بگیریم. در غیر این صورت، تاثیرات واقعی دارو در لابلای تغییرات طبیعی فشارخون مخفی خواهد ماند. فکر کنید که اگر واقعاً میزان ضریب هوشی همه مردان و همه زنان صد در صد مساوی باشد (یعنی انحراف معیار ضریب هوشی در هر جنس صفر باشد) برای مشخص نمودن تفاوت این متغیر در دو جنس شما فقط به سنجش ضریب هوشی در یک مرد و یک زن نیاز دارید! حال اگر میزان پراکندگی طبیعی ضریب هوشی در هر دو جنس بسیار زیاد باشد، یعنی در هر دو جنس افرادی با ضریب هوشی بسیار بالا و بسیار پایین داشته باشید، برای یافتن تفاوت احتمالی در میانگین ضریب هوشی زنان و مردان باید حجم نمونه بالایی داشته باشیم. پس هر چه متغیر یا متغیرهای مورد نظر در جامعه بیشتر پراکنده باشند، حجم نمونه مورد نیاز برای پاسخ به سوالات پژوهش را افزایش می‌دهند. البته مفهوم این پراکندگی در خصوص متغیرهای غیر عددی فرق می‌کند. همانگونه که می‌دانید، انحراف استاندارد یک متغیر دو حالتی بر اساس توزیع دو جمله‌ای از طریق فرمول زیر محاسبه می‌شود

$$SD = \sqrt{p * (1 - p)}$$

یعنی برای محاسبه انحراف معیار یک متغیر دو حالتی مانند احتمال بیمار شدن، باید این احتمال را در احتمال بیمار نشدن ضرب نمود و سپس جذر آن را حساب کرد. بر این اساس اگر میزان فراوانی پدیده (p) به سمت صفر یا یک میل نماید مقدار انحراف معیار کم خواهد شد و به طبع آن برای پاسخ به سوال پژوهش باید حجم نمونه کمتری در نظر گرفت. احتمالاً این جمله را بارها شنیده‌اید که در زمانی که تخمینی از شیوع یک پدیده ندارید برای تعیین حجم نمونه مقدار شیوع را ۵۰٪ بگیرید تا بیشترین حجم نمونه محاسبه گردد و به عبارتی روش محافظه‌کارانه را در پیش بگیرید که با توضیح فوق نیز همخوانی دارد؛ زیرا در این حالت انحراف معیار حداکثر خواهد شد. ولی حال می‌خواهم موضوع را پیچیده‌تر نمایم. مطالعه بر روی پدیده‌های نادر احتمالاً سخت‌تر است. مثلاً برای تخمین میزان بروز کم‌کاری مادرزادی تیروئید که شیوع آن حدود ۱ در هر ۱۰۰۰ نوزاد و یا نزدیک به آن است قطعاً تعداد نمونه‌های بیشتری نیاز است تا مطالعه بر روی موضوعی مانند نوزاد کم‌وزن که شیوع بسیار

بزرگتر از حجم نمونه مورد نیاز است.

محاسبه حجم نمونه برای مطالعات توصیفی

در مطالعات توصیفی، محقق به دنبال برآورد یک پارامتر مشخص است که در اکثر اوقات این پارامتر یک میانگین و یا یک نسبت است. به عنوان مثال تصور نمایید که هدف ما برآورد میانگین وزن نوزادان در بدو تولد در شهر کرمان است و یا به دنبال تخمین درصد نوزادان با وزن کمتر از ۲۵۰۰ گرم در این شهر هستیم. فرمول محاسبه در این دو حالت بسیار نزدیک به هم است که در ادامه شرح داده خواهد شد. قبل از آن بهتر است توضیحات کمی داده شود تا بهتر بتوانید جزئیات این فرمول‌ها را درک نمایید.

حتماً به خاطر دارید که برای برآورد یک میانگین و یا یک نسبت، چند نکته بسیار مهم هستند. نکته اول ضریب اطمینان است و نکته دوم دامنه اطمینان. به عنوان مثال فرض کنید که در یک مطالعه بر روی ۱۰۰ نوزاد کرمانی، میانگین وزن ۲۹۰۰ گرم محاسبه شده است. این عدد آماره است چرا که در یک نمونه بدست آمده و ما معمولاً سعی می‌کنیم بر اساس آماره، مقدار پارامتر یعنی میانگین وزن نوزادان در کل جامعه را برآورد کنیم. مثلاً بگوییم که اگر وزن در کل جامعه (در کل نوزادان متولد شده در شهر کرمان) دقیقاً برابر ۲۹۰۰ گرم نیست، لیکن به احتمال ۹۵٪ عددی بین ۲۷۰۰ تا ۳۱۰۰ گرم است (2900 ± 200). فاصله بین ۲۷۰۰ تا ۳۱۰۰ گرم، دامنه اطمینان و عدد ۹۵٪ ضریب اطمینان را نشان می‌دهد. برای محاسبه حجم نمونه باید تعیین نماییم که برای برآورد مورد نظر چه ضریب اطمینان و چه دامنه اطمینانی مد نظر است و طبیعی است که هر چه دامنه اطمینان باریکتر (یعنی دقت بیشتر) و ضریب اطمینان بیشتری (یعنی خطای کمتری) را بخواهیم، باید حجم نمونه بزرگتری را نیز اختیار نماییم.

در برآورد یک نسبت نیز همین مفاهیم وجود دارد. مثلاً اگر در نمونه، شیوع کم وزنی ۱۵٪ در آمده باشد، ممکن است برآورد شیوع کم‌وزنی در کل جامعه با اطمینان ۹۵٪ برابر 15 ± 5 ٪ شود. ولی در بحث حجم نمونه مسیر حرکت برعکس است. می‌گوییم اگر شیوع حدود ۱۵٪ باشد و بخواهیم با اطمینان ۹۵٪ و حداکثر خطای ± 5 ٪ شیوع در کل جامعه را برآورد نماییم، باید چه تعداد نمونه بگیریم؟

با توجه به توضیحات فوق، در قدم اول برای محاسبه حجم نمونه باید بر اساس جدول z مقدار این عدد را برای ضریب اطمینان

مورد نظر پیدا نماییم. حتماً به یاد دارید که مقدار عدد z برای دامنه اطمینان ۹۵٪ برابر ۱/۹۶ است که در محاسبات اکثراً برابر ۲ در نظر گرفته می‌شود. سپس برای وارد نمودن مفهوم دامنه اطمینان باید عدد d را محاسبه نمود؛ این عدد برابر با حداکثر فاصله بین برآورد نقطه‌ای با دامنه پایین و یا دامنه بالای اطمینان است. در مثال‌های فوق مقدار عدد d برای برآورد میانگین وزن ۲۰۰ گرم و برای برآورد نسبت نوزادان کم‌وزن ۵٪ می‌باشد.

آخرین جز برای محاسبه حجم نمونه برآورد انحراف معیار متغیر مورد نظر است. شاید سخت‌ترین قسمت نیز همین برآورد باشد. برای متغیرهای عددی انحراف معیار را باید بر اساس اطلاعات مطالعات مشابه و یا بر اساس اطلاعات یک پیش مطالعه تخمین زد؛ ولی در عمل و در بسیاری از مواقع امکان دستیابی به اطلاعات چنین پژوهش‌هایی مقدور نیست. اما یک راه میانبر وجود دارد. فکر کنید که احتمالاً حداقل و حداکثر مقدار متغیر در جامعه مورد مطالعه چقدر است. سپس دامنه تغییرات را محاسبه و آن را تقسیم بر ۶ نمایید. به عنوان مثال اگر فکر نماییم که کمترین و بیشترین وزن نوزادان در کرمان ۱۵۰۰ و ۴۵۰۰ گرم است، دامنه تغییرات برابر ۳۰۰۰ گرم خواهد بود و یک ششم این دامنه ۵۰۰ گرم است که می‌تواند تخمینی از انحراف معیار متغیر وزن در جامعه باشد. آیا می‌توانید منطق چنین برآوردی از انحراف معیار را بیان فرمایید؟ سوال چندان سخت نیست. به یاد بیاورید که تقریباً میانگین به اضافه و منهای ۳ انحراف معیار حدوداً ۹۹/۹٪ از جامعه را پوشش می‌دهد. حال با این راهنمایی باید حدس بزنید که چرا دامنه تغییرات تقسیم بر ۶ تخمینی از انحراف معیار خواهد بود (۴،۲-۶).

اما داستان انحراف معیار در برآورد یک نسبت، موضوعی کاملاً متفاوت است. برای تخمین این شاخص باید برآوردی از میزان شیوع متغیر مورد نظر داشت. همانگونه که قبلاً نیز شرح داده شده بود، انحراف معیار در برآورد یک نسبت، بر اساس توزیع دو جمله‌ای محاسبه می‌شود و برابر جذر شیوع پدیده مورد نظر ضربدر، یک منهای این شیوع است. لذا برای محاسبه حجم نمونه باید یک تخمین از شیوع متغیر داشت. مثلاً برای محاسبه شیوع وزن کمتر از ۲۵۰۰ گرم در نوزادان باید تخمینی از این متغیر در اختیار داشت. گاه این تخمین غیر ممکن است. در این موارد بیان می‌شود که بهتر است شیوع ۵۰٪ (یا به عبارتی ۰/۵) در نظر گرفته شود. البته با فرض ثابت بودن مقدار d این فرض حجم نمونه را به صورت محافظه‌کارانه محاسبه می‌کند. یعنی عددی را بیان می‌کند که قطعاً شما را به نتیجه خواهد رساند؛ ولی البته

ممکن است بیش از حد نیاز باشد.

خوب با این توضیحات سراغ فرمول‌های محاسبه حجم نمونه برای برآورد یک میانگین و یا یک نسبت می‌رویم. فرمول‌های زیر را به دقت ملاحظه فرمایید و سعی کنید اجزای آن را با توجه به توضیحات چند پاراگراف بالا به صورت عمیق درک نمایید.

$$n = \frac{z^2 * SD^2}{d^2} \quad \text{برآورد میانگین}$$

$$n = \frac{z^2 * p(1-p)}{d^2} \quad \text{برآورد نسبت}$$

اگر فرمول‌های فوق را درک نموده‌اید، محاسبه نمایید که برای تخمین میانگین وزن نوزادان کرمانی در صورتی که انحراف معیار وزنشان ۵۰۰ گرم باشد و بخواهیم با دقت ۰.۹۵٪ و حداکثر با خطای ۱۰۰ گرم (d) آن را برآورد نماییم، چه تعداد نمونه لازم خواهیم داشت؟

اگر درست عمل نمایید باید به عدد تقریباً ۱۰۰ نفر رسیده باشید.

حال از فرمول دوم استفاده نمایید و محاسبه نمایید اگر تخمین اولیه ما از کم‌وزنی حدود ۰.۱۵٪ باشد و بخواهیم با اطمینان ۰.۹۵٪ و حداکثر خطای ۰.۰۵٪ آن را برآورد نماییم به چند نمونه نیاز داریم؟ بله درست است تقریباً ۲۰۰ نمونه مورد نیاز است. اما یک اشتباه رایج که ممکن است در این قسمت صورت پذیرد آن است که به جای نسبت (۰/۱۵)، درصد (۰/۱۵٪) را در فرمول وارد نمایید. توجه داشته باشید که اگر بخواهید درصد را بکار برید، حتماً در داخل پرانتز بجای (1-p) باید (100-p) را قرار دهید.

حال که مسیر محاسبه را درک نمودید، بگذارید یک روش میانبر را خدمتان بیان کنیم. همانگونه که شرح داده شد، برآورد انحراف معیار و یا تخمین احتمالی نسبت مورد مطالعه برای تخمین حجم نمونه، سخت‌ترین پیش‌فرض است. برای آسان‌تر شدن کار می‌توان میزان اثر مورد نظر را کسری از انحراف معیار و یا کسری از فراوانی متغیر مورد نظر دانست و به این شکل صورت و مخرج کسر را آسان ساخت. به عنوان مثال فکر کنید که می‌خواهیم وزن نوزادان را محاسبه نماییم ولی هیچ تخمینی از انحراف معیار نداریم. برای استفاده از این روش میانبر فرض می‌کنیم که مثلاً d یک‌دهم مقدار انحراف معیار باشد؛ در این صورت فرمول به شکل زیر ساده خواهد شد:

$$n = \frac{z^2 * SD^2}{(0.1SD)^2} = n = \frac{z^2}{(0.1)^2} = z^2 * 100$$

مشابه همین روش را می‌توان در برآورد یک نسبت استفاده نمود؛ مثلاً می‌توان فرض نمود که مقدار d برابر ۰/۲ فراوانی متغیر

مورد نظر باشد. به این شکل فرمول اینچنین خواهد شد:

$$n = \frac{z^2 * p(1-p)}{(0.2p)^2} = \frac{z^2 * (1-p)}{(0.04p)} = 25z^2 * \frac{1-p}{p} = 25z^2 * \left(\frac{1}{p} - 1\right)$$

اگر خوب به این دو فرمول جدید نگاه نمایید، متوجه خواهید شد در زمانی که مقدار انحراف معیار کوچک باشد و یا شیوع متغیر مورد نظر کم باشد و بخواهیم d، کسری از انحراف معیار و یا شیوع باشد، حجم نمونه افزایش می‌یابد که تا حدودی برخلاف توضیحات قبلی است. البته در صورت درک عمیق، متوجه خواهید شد که دلیل آن چیست. واقعیت این گونه است، مثلاً وقتی که بخواهیم یک متغیر با شیوع ۰.۳۰٪ را برآورد نماییم، میزان d برابر ۰.۳٪ یعنی برآورد دامنه اطمینان ۰.۳٪±۰.۳۰٪ چندان غیر دقیق نخواهد بود. ولی اگر میزان شیوع ۰.۳٪ باشد، آنوقت میزان d برابر ۰.۳٪ یعنی دامنه اطمینان ۰.۳٪±۰.۳٪ زیاد دقیق به نظر نمی‌رسد و در این صورت برای دقیق عمل نمودن باید مقدار d را بسیار کوچکتر فرض نمود. طبیعی است که در صورت ثابت فرض نمودن مقدار d، یعنی ۰.۳٪، هرچه مقدار p به عدد ۰.۵۰٪ نزدیک شود مقدار حجم نمونه افزایش می‌یابد؛ ولی اگر مقدار d را کسری از p بگیریم، با کوچکتر شدن مقدار p و به طبع آن کوچکتر شدن d، حجم نمونه افزایش می‌یابد ۲-۴.

اما حتماً یک سوال برای شما به وجود آمده و آن این است که نسبت بین d و انحراف معیار و یا شیوع چقدر باید باشد. به خاطر دارید که در مثال‌های بالا این نسبت ۰/۱ و ۰/۲ فرض شد. بیان می‌شود که این نسبت را باید بسته به شرایط و موضوع تحقیق و همچنین مطالعات مشابه انتخاب نمود؛ ولی آنچه که بیشتر عرف است و به آن عمل می‌شود آن است که نسبت مذکور را معمولاً بین ۰/۱ تا ۰/۲۵ در نظر می‌گیرند و روشن است که هرچه نسبت مذکور به ۰/۱ نزدیکتر شود، مقدار d کوچکتر می‌شود و لذا حجم نمونه محاسبه شده بیشتر خواهد شد.

محاسبه حجم نمونه برای مقایسه دو نسبت

بعضی اعتقاد دارند که مصرف آهن در دوران بارداری ممکن است شانس بروز زردی نوزادی را افزایش دهد. در یک تحقیق بر آن شدیم که نشان دهیم آیا میزان بروز زردی نوزادی در فرزندان که مادرانشان در هنگام بارداری قرص آهن مصرف نموده‌اند (گروه ۱) با آنهایی که قرص آهن مصرف نموده بودند (گروه ۲) برابر است یا خیر؟ حال سوال مهم آن است که از هر گروه چند نفر باید وارد مطالعه شوند؟

برای پاسخ به این سوال باید بعضی نکات روشن شود. اولاً باید بگوییم که حداقل تفاوت بالینی مهم چه میزان است؟ آیا اگر واقعاً

چیست؟ بله درست حدس زده‌اید:

$$\text{Power} = 1 - \beta$$

برگردیم به سوال اصلی یعنی تخمین حجم نمونه در مقایسه دو نسبت. زمانی که می‌خواهیم حجم نمونه را تعیین نماییم باید بیان کنیم که حداکثر خطای نوع اول و دوم مورد قبول چه میزان است. طبیعی است که هرچه بخواهیم دقیق‌تر کار نماییم، یعنی خطاهای نوع اول و دوم را کاهش دهیم باید حجم نمونه بیشتری داشته باشیم. ولی به صورت کلی معمولاً خطای نوع اول را ۰.۰۵٪ و خطای نوع دوم را ۰.۱۰٪ و در بعضی موارد ۰.۲۰٪ در نظر می‌گیریم. خوب با این توضیحات، فرمول محاسبه حجم نمونه برای مقایسه دو نسبت را مرور کنیم:

$$n = \frac{\left[z_{(1-\frac{\alpha}{2})} \sqrt{2\bar{p}(1-\bar{p})} + z_{1-\beta} \sqrt{p_0(1-p_0) + p_1(1-p_1)} \right]^2}{\bar{p}^2}$$

$$\bar{p} = \frac{p_1 + p_0}{2}$$

فرمول‌ها در ظاهر پیچیده هستند، اما اگر دقت نمایید خواهید دید که علیرغم طولانی بودن فرمول، استفاده از آن چندان دشوار نیست. مقدار p_1 و p_0 برابر با نسبت در دو گروه مورد مقایسه است. در تمرین فوق یعنی مقدار تخمین احتمالی میزان بروز زردی در گروه اول و دوم و طبیعی است که مقدار تفاوت این دو کسر برابر با حداقل تفاوت مهمی است که به دنبال کشف آن هستیم. مثلاً فکر کنید احتمالاً میزان بروز زردی در گروه اول که مادرانشان قرص آهن مصرف نموده‌اند ۰.۱۵٪ است و فکر می‌کنید در صورتی اثر مصرف قرص آهن بر روی زردی نوزادی مهم است که شیوع در گروه دوم حداقل ۰.۲۰٪ باشد، یعنی به دنبال یک اثر حداقلی ۰.۰۵٪ برای مصرف قرص آهن هستیم. بر اساس فرمول کوچکتر مشخص است که مقدار \bar{p} برابر میانگین ساده این دو نسبت یعنی ۰.۱۷۵٪ خواهد بود.

اما مقدار اعداد z را باید بر اساس جدول مربوطه و متناسب با خطاهای نوع اول و دوم محاسبه نماییم. اگر خطای نوع اول را

مثلاً یک دهم درصد تفاوت در میزان بروز در دو گروه وجود داشته باشد از نظر شما مهم است؟ طبیعتاً هرچه میزان تفاوت مهم بالینی کمتر در نظر گرفته شود، حجم نمونه بیشتر خواهد بود. درست مثل حالتی که بخواهیم اشیاء ریزتری را ببینیم و لذا به ذره‌بین قوی‌ترین نیاز داریم. موضوع بعدی که باید مد نظر باشد آن است که میزان خطاهای قابل قبول نوع اول و دوم آماری چقدر است؟ آیا شما با این دو نوع خطای آماری آشنا هستید؟

خطای نوع اول یعنی رد نمودن یک فرضیه صفر درست و احتمال بروز چنین خطایی را با α نشان می‌دهیم؛ و خطای نوع دوم یعنی قبول نمودن یک فرضیه صفر غلط و احتمال بروز چنین خطایی را با β نشان می‌دهیم. آیا درک این جملات دشوار است؟ با توضیحات بعدی موضوع روشن‌تر خواهد شد (۴،۷).

در مثال بالا، فرضیه صفر بیان می‌کند که میزان بروز زردی نوزادی در دو گروه برابر است. خوب اگر این فرضیه واقعاً درست باشد، آیا ممکن است به اشتباه نتیجه‌گیری نماییم که میزان بروز زردی در این دو گروه تفاوت دارد؟ بله این احتمال وجود دارد. اگر به اشتباه در حالی که میزان بروز زردی در دو گروه برابر است نتیجه‌گیری نماییم که این تفاوت وجود دارد، مرتکب خطای نوع اول شده‌ایم.

در مقابل اگر واقعاً فرضیه صفر غلط باشد و میزان بروز زردی در دو گروه متفاوت باشد، آیا ما می‌توانیم مطمئن باشیم که حتماً و قطعاً این تفاوت را در تحقیق خود کشف خواهیم نمود؟ مسلماً پاسخ منفی است. ما با یک قدرت و توانی می‌توانیم تفاوت را کشف نماییم که ۰.۱۰۰٪ نیست. احتمال عدم کشف نمودن تفاوت واقعی را خطای نوع دوم می‌نامیم. در اینجا باید با یک اصطلاح دیگر نیز آشنا شویم و آن توان (power) آزمون آماری است و بیان می‌کند که در صورت وجود تفاوت (عدم صحت فرضیه صفر) با چه قدرت و دقتی می‌توانیم این تفاوت را کشف نماییم. پس توان آزمون نشان می‌دهد که قدرت یافتن ارتباطات و یا اختلافات واقعی چه میزان است؛ درست مثل قدرت یک ذره‌بین در دیدن اشیاء.

نتیجه‌گیری نهایی

وضعیت واقعی فرضیه صفر که در عمل ما از آن بی‌اطلاع هستیم و در پی کشف آن	قبول فرضیه صفر یعنی بیان عدم ارتباط	رد فرضیه صفر یعنی بیان ارتباط و یا اختلاف بین متغیرها
درست است یعنی اختلاف و یا ارتباط واقعی بین متغیرها وجود ندارد	۱- α	؟
غلط است یعنی اختلاف و یا ارتباط واقعی بین متغیرها وجود دارد	؟	power

برای روشن شدن مطالب بالا لطفاً دو خانه خالی جدول زیر را پر کنید. حالاً آیا می‌توانید رابطه بین توان آزمون و خطای نوع دوم

برای روشن شدن مطالب بالا لطفاً دو خانه خالی جدول زیر را پر کنید.

حالاً آیا می‌توانید رابطه بین توان آزمون و خطای نوع دوم

تصادفی ساده، تعدیل حجم نمونه برای آزمون‌های غیرپارامتریک و همچنین حجم نمونه در مطالعات کیفی مطالبی آورده خواهد شد.

منابع

1. Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. *Journal of American Medical Association* 1994; 272: 122-4.
2. Georgiadis MP, Johnson WO, Gardner IA. Sample size determination for estimation of the accuracy of two conditionally independent tests in the absence of a gold standard. *Preventive Veterinary Medicine*, 2005; 71: 1-10.
3. Jones SR, Carley S, Harrison M, An introduction to power and sample size estimation. *Emergency Medical Journal* 2003; 20: 453-8.
4. Delucchi KL. Sample size estimation in research with dependent measures and dichotomous outcomes. *American Journal of Public Health* 2004; 94: 372-7.
5. Eng J. Sample size estimation: how many individuals should be studied? *Radiology* 2003; 227: 309-13.
6. Lerman J. Study design in clinical research: sample size estimation and power analysis. *Can J Anaesth*, 1996. 43: 184-91.
7. Lerman J. Sample size estimation for nominal data. *Canadian Journal of Anaesthesia*, 1997; 44: 901.
8. Houle TT, Penzien DB, Houle CK. Statistical power and sample size estimation for headache research: an overview and power calculation tools. *Headache*, 2005; 45: 414-8.
9. Eng J. Sample size estimation: a glimpse beyond simple formulas. *Radiology* 2004; 230: 606-12.

محاسبه حجم نمونه برای مقایسه دو میانگین

حال با توجه به توضیحات فوق، استفاده از فرمول مقایسه دو میانگین بسیار ساده خواهد بود؛ چون با مفهوم خطاهای نوع اول و دوم و نقش حداقل تفاوت مهم آشنا شده‌اید. فرمول محاسبه در این حالت به شرح زیر است:

$$n = \frac{2(z_{1-\frac{\alpha}{2}} + z_{1-\beta})^2 \sigma^2}{d^2}$$

مثلاً اگر بخواهیم تفاوت بین غلظت بیلی‌روبین نوزادان در دو گروه را با حداکثر خطاهای نوع اول و دوم ۰/۵٪ و ۱۰٪ بسنجیم و تخمین ما از انحراف معیار بیلی‌روبین در خون نوزادان حدود ۴ باشد و به دنبال کشف حداقل ۲ واحد تفاوت (d) باشیم، حجم نمونه حدوداً ۸۵ نوزاد در هر گروه خواهد بود. در این مثال حجم نمونه کم است و دلیل آن زیاد بودن نسبت بین میزان اثر (۲) و انحراف معیار (۴) است. در توضیحات فوق بیان شد که معمولاً این نسبت حداکثر ۰/۲۵ در نظر گرفته می‌شود؛ ولی در حالت فوق این نسبت ۰/۵ است (۹-۴،۷).

در ادامه و در شماره‌های بعد، شیوه تعدیل حجم نمونه در شرایط خاص مانند مقایسه بیش از ۲ گروه، نمونه‌گیری غیر