

تعیین حجم نمونه در مطالعات جور شده

سیدمهدی سادات‌هاشمی^{*} (Ph.D)، راهب قربانی^۱ (Ph.D)، بهروز کاوه‌یی^۲ (Ph.D)

۱- دانشگاه علوم پزشکی سمنان، دانشکده پزشکی، گروه پزشکی اجتماعی

۲- وزارت علوم، تحقیقات و فن‌آوری

چکیده

سابقه و هدف: با داده‌های جور شده هنگامی مواجه می‌شویم که هدف مطالعه نشان دادن تأثیر یک عامل خاص بر روی یک نتیجه معین باشد و در آن برای بررسی تأثیر اصلی آن عامل، نمونه‌ها به لحاظ سایر عوامل با یکدیگر یکسان یا جور شده باشند. در مطالعات مذکور، پاسخ می‌تواند از هر نوع (اسمی، رتبه‌ای، فاصله‌ای یا نسبتی) باشد. از آنجایی که مطالعات مذکور به‌طور گسترده‌ای در علوم پزشکی انجام می‌شود و تعداد نمونه به کارگرفته شده در آن‌ها بایستی کفایت لازم برای اجرای آزمون‌های معنی‌داری را داشته باشد، لذا بر آن شدیم تا با در نظر گرفتن ماهیت پاسخ در چنین مطالعاتی، روابط و روش‌های مناسب برای تعیین حجم نمونه را ارائه کرده و نتایج را به مطالعات مورد-شاهدی جور شده با چند شاهد برای یک مورد تعمیم دهیم. هم‌چنین برای نشان دادن به کارگیری عملی روابط ریاضی از مثال‌های عملی در حیطه پزشکی استفاده نموده‌ایم.

مواد و روش‌ها: در این قسمت روابط مربوط به تعیین حجم نمونه در حالت‌های داده‌های زوجی با پاسخ‌های کمی (فاصله‌ای یا نسبتی)، داده‌های زوجی رتبه‌ای و داده‌های زوجی اسمی را معرفی و تشریح نموده‌ایم. یافته‌ها: به منظور درک بیش‌تر خوانندگان با استفاده از مثال‌های کاربردی که از مطالعات واقعی در زمینه علم پزشکی برگرفته شده‌اند، نحوه استفاده از روابط تعیین حجم نمونه در مطالعات زوجی را تشریح نموده‌ایم. نتیجه‌گیری: چنان‌چه طراحی یک مطالعه بر مبنای مقایسه‌های زوجی جور شده باشد، تعیین حجم نمونه با استفاده از روابط معرفی شده در این مقاله، توان آزمون‌ها و صحت نتایج به دست آمده را افزایش خواهد داد.

واژه‌های کلیدی: حجم نمونه، مطالعات جور شده، مطالعات مورد-شاهدی، مطالعات قبل و بعد، مطالعات تقاطعی

مقدمه

یکی از مهم‌ترین مواردی که بایستی در یک پژوهش در نظر گرفته شود تعداد نمونه است. زیرا اجرای یک آزمون معنی‌دار آماری در یک مطالعه تحلیلی و یا برآورد دقیقی از یک پارامتر در یک مطالعه توصیفی نیازمند به‌کارگیری تعداد مناسبی از نمونه می‌باشد. کم‌تر بودن تعداد، از آن‌چه لازم است منجر به اریب شدن آزمون‌ها و افت توان آن‌ها و بیش‌تر بودن آن منجر به مشکلاتی از قبیل افزایش هزینه، زمان و در برخی از مطالعات بالینی، که در حیطه علوم پزشکی قرار می‌گیرند،

منجر به در مخاطره قرار گرفتن تعداد افراد بیش‌تری می‌شود. در همه انواع مطالعات کمی، عوامل گوناگونی بر انتخاب مناسب تعداد نمونه دخالت دارند که عبارتند از ۱- پراکندگی (واریانس) صفت مورد بررسی، ۲- توقع پژوهش‌گر از نتایج آماری (این قسمت در قالب خطای نوع اول و دوم قرار می‌گیرد که خواننده علاقه‌مند می‌تواند جزئیات مربوط به آن‌ها را در منبع [۱] ملاحظه نماید)، ۳- در مطالعات مقایسه‌ای در نظر گرفتن مقدار اختلافی که باید آشکار شود. مطالعات جور شده زوجی، منجر به پدید آمدن داده‌های

* نویسنده مسئول. تلفن: ۰۲۳۱-۳۳۳۲۹۱۲، نمابر: ۰۲۳۱-۳۳۳۲۹۱۲، E-mail: sadat-hashemi@sem-ums.ac.ir

(۱) حجم نمونه برای داده‌های زوجی با پاسخ‌های کمی (فاصله‌ای یا نسبتی).

۱-۱) آزمون‌های پارامتری. فرض کنید Y_1 و Y_2 به ترتیب پاسخ‌های زوج مورد-شاهد باشند. تفاوت بین زوج‌ها که عبارت است از $\delta = Y_1 - Y_2$ ، از لحاظ بالینی مورد توجه می‌باشد (به عنوان مثال در یک مطالعه برای بررسی تأثیر یک داروی خاص در کاهش کلسترول خون، Y_1 اندازه کلسترول فرد قبل از مطالعه و Y_2 اندازه کلسترول خون فرد پس از مصرف دارو می‌باشد، در این صورت δ تفاوت کلسترول قبل و بعد از درمان است). حداقل حجم نمونه لازم برای آشکار کردن این تفاوت به کمک یک آزمون معنی‌داری دوطرفه در سطح معنی‌داری α و توان $1 - \beta$ از رابطه زیر به دست می‌آید [۱۰]:

$$n = \frac{\left(Z_{1-\frac{\alpha}{2}} + Z_{1-\beta} \right)^2 \sigma_{\delta}^2}{\delta^2} + \frac{Z_{1-\frac{\alpha}{2}}^2}{2}$$

در این رابطه σ_{δ}^2 عبارت است از واریانس $Y_1 - Y_2$ و $Z_{1-\frac{\alpha}{2}}$ و $Z_{1-\beta}$ صدک‌های توزیع نرمال استاندارد و متناظر با $\frac{\alpha}{2}$ و β هستند. هم‌چنین دومین قسمت معادله (۱)، عامل تصحیحی است که استفاده از آزمون t زوجی را برای بررسی اختلاف معنی‌دار Y_1 و Y_2 مجاز می‌کند. زیرا با تقریب خیلی خوبی سطح معنی‌داری α را برای اجرای این آزمون ثابت نگه‌می‌دارد [۱۱]. اگر واریانس بین موارد (یا بین شاهد‌ها) σ^2 باشد، آن‌گاه $\sigma_{\delta}^2 = 2\sigma^2(1 - \rho)$ که در آن ρ ضریب هم‌بستگی پیرسون بین Y_1 و Y_2 است [۱۲]. فریسان و یوکاک استفاده از معادله (۱) را هنگامی مجاز می‌دانند که ρ در بازه عددی (۰/۷۵، -۰/۷۵) قرار داشته باشد و برای سایر مقادیر که خارج از این بازه هستند، استفاده از تحلیل حساسیت بر روی حجم نمونه را پیشنهاد نموده‌اند [۱۳].

زوجی می‌شوند و هدف آن‌ها نشان دادن تأثیر یک عامل خاص بر روی یک نتیجه معین می‌باشد و در آن‌ها برای بررسی تأثیر اصلی آن عامل، نمونه‌ها به لحاظ سایر عواملی که به آن‌ها محدودش‌گر گفته می‌شود با یک‌دیگر یک‌سان یا جورشده باشند. این کار بنا به دلایل گوناگون می‌تواند باعث کاهش یافتن تعداد نمونه نسبت به حالتی شود که همان مطالعه با روش غیرزوجی نیز قابل اجراء است [۲،۳]. در حیطه تحقیقات پزشکی و بهداشتی یکی از مهم‌ترین کاربردهای یک مطالعه زوجی، کنترل عوامل محدودش‌کننده می‌باشد. کارآزمایی‌های بالینی اعم از تقاطعی (Cross-over)، قبل و بعد و نیز مطالعات مورد-شاهد در این حیطه قرار می‌گیرند [۴،۷]. در همه آن‌ها هر مورد با شاهده‌ی که به لحاظ عوامل محدودش‌گری چون سن، جنس و... (که در مطالعه اثر آن‌ها بر نتیجه مورد توجه نمی‌باشد) با وی یک‌سان است، جور شده و تحت بررسی قرار می‌گیرد. به عنوان مثال در کارآزمایی‌های بالینی به روش تقاطعی و یا روش قبل و بعد، شاهد جور شده عبارت است از خود فرد در مرحله بعدی.

در مطالعات مذکور، پاسخ می‌تواند از هر نوع (اسمی، رتبه‌ای، فاصله‌ای یا نسبتی) باشد. در حالتی که پاسخ از نوع کیفی (اسمی یا رتبه‌ای) باشند، چنان‌چه در یک زوج، پاسخ مورد با شاهد یک‌سان باشد به آن زوج سازگار (Concordant) و در غیر این صورت ناسازگار (Discordant) گفته می‌شود [۸،۹].

از آن جایی که مطالعات مذکور به طور گسترده‌ای در علوم پزشکی انجام می‌شود و تعداد نمونه به کارگرفته شده در آن‌ها بایستی کفایت لازم برای اجرای آزمون‌های معنی‌داری را داشته باشد، لذا بر آن شدیم تا با در نظر گرفتن ماهیت پاسخ در چنین مطالعاتی، روابط و روابط مناسب برای تعیین حجم نمونه را ارائه کرده و نتایج را به مطالعات مورد-شاهدی جورشده با چند شاهد برای یک مورد تعمیم دهیم. هم‌چنین برای نشان دادن به کارگیری عملی روابط ریاضی از مثال‌های عملی در حیطه پزشکی استفاده نموده‌ایم.

تعداد زوج‌های مورد انتظار برای تفاضل d_i به کل زوج‌ها باشد، آن‌گاه مقدار مورد انتظار برای تفاوت کل رتبه‌ها یعنی $\mu = Y_1 - Y_2$ و واریانس آن یعنی σ^2 عبارت است از: (۳)

$$\mu = \sum_{j=1}^{2(K-1)} d_j \pi_j \quad \sigma^2 = \sum_{j=1}^{2(K-1)} d_j^2 \pi_j - \mu^2$$

که در آن‌ها $2(K-1)$ تعداد کل طبقات ناجور می‌باشد. از جای‌گذاری μ و σ^2 در روابط (۳) به جای δ و σ_δ^2 در

معادله (۱) می‌توان تعداد زوج‌های ناجور با n_d را به دست آورد. چنان‌چه مقیاس دارای K سطح باشد، احتمال این که

یک زوج به صورت جور باشد، عبارت است از $\frac{K}{K^2} = \frac{1}{K}$

لذا احتمال ناجور بودن آن عبارت است از $1 - \frac{1}{K} = \frac{K-1}{K}$

پس تعداد کل زوج‌های لازم عبارت است از: (۴)

$$n = \frac{n_d}{\frac{K-1}{K}} = \frac{Kn_d}{K-1}$$

(۳) حجم نمونه برای داده‌های زوجی اسمی.

(۱-۳) داده‌های زوجی دو حالته. برای تعیین حجم نمونه در این گونه حالات، نسبت زوج‌های ناسازگار در نظر گرفته می‌شود و استفاده از آزمون مک‌نمار مناسب است [۱۴، ۱۵].

برای یک مطالعه زوجی مورد-شاهدی نمادگذاری طبق جدول ۱ می‌باشد. این نمادگذاری با اندکی تغییر برای مطالعات تقاطعی و قبل و بعد نیز قابل استفاده است.

جدول ۱. نمادگذاری نتایج یک مطالعه مورد-شاهدی زوجی با پاسخ

دو حالته

		شاهد
		مورد
با عامل مواجهه ندارد	با عامل مواجهه دارد	با عامل مواجهه دارد
π_{10}	π_{00}	با عامل مواجهه ندارد
π_{11}	π_{01}	

در این جدول، نسبت‌های مربوط به زوج‌های سازگار عبارت است از: π_{11} و π_{00} و نسبت‌های مربوط به زوج‌های ناسازگار عبارت است از: π_{10} و π_{01} . هرگاه

(۲-۱) آزمون ناپارامتری ویلکاکسون. هرگاه به جای استفاده از آزمون‌های پارامتری نظیر Z یا t استفاده از آزمون ناپارامتری ویلکاکسون مدنظر محقق باشد، در صورتی که داده‌ها کمی باشند، نوتر روشی را برای تعیین حجم نمونه پیشنهاد نموده است [۱۴] که جزئیات آن به شرح زیر می‌باشد:

اگر d_i و d_j تفاضل پاسخ‌های دو زوج i و j باشند که به طور تصادفی از جامعه زوج‌ها انتخاب شده‌اند، آن‌گاه فرض کنید P_0 احتمال این باشد که $d_i + d_j > 0$ ؛ بنابراین

تحت فرض H_0 که تأکید بر عدم تفاوت زوج‌ها، یعنی صفر بودن تفاضل آن‌ها دارد، خواهیم داشت $P_0 = 0.5$. حال

فرض کنید که پژوهش‌گر برای فرض مقابل مقدار احتمال P_1 را کاندید کرده باشد، در این صورت حداقل حجم نمونه لازم برای اجرای آزمون ناپارامتری دو طرفه رتبه علامت ویلکاکسون در سطح معنی‌داری α و توان $1 - \beta$ عبارت است از: (۲)

$$n = \frac{\left(Z_{1-\frac{\alpha}{2}} + Z_{1-\beta} \right)^2}{3(P_1 - 0.5)^2}$$

مشکل این روش در غیرمجاز بودن هم‌رتبه‌ها در آن است. لذا چنان‌چه برای متغیرهای رتبه‌ای از آن استفاده شود با کاهش تعداد طبقات متغیر و یا افزایش تعداد هم‌رتبه‌ها، دقت آن رو به کاهش می‌گذارد.

(۲) حجم نمونه برای داده‌های زوجی رتبه‌ای.

فرض کنید متغیر پاسخ رتبه‌ای دارای $K (K > 2)$ سطح باشد و Y_{i1} و Y_{i2} رتبه‌های i امین زوج و $d_i = Y_{i1} - Y_{i2}$

تفاضل رتبه‌های این زوج باشد. بنابراین امکان $2(K-1)$ تعداد زوج با تفاضل غیرصفر، که در تعیین حجم نمونه نقش

اصلی را ایفا می‌کنند، برای d_i وجود دارد. برای مثال در یک مقیاس سه سطحی، امکان چهار اختلاف غیرصفر وجود دارد

که عبارتند از ۲، -۱، -۲، ۱ و ۲. حال فرض کنید که π_i نسبت

هر طبقه را تعیین نمود که عبارتند از: $\pi_{01} = \frac{\psi}{\psi+1}$ و

$$\pi_{10} = \frac{1}{\psi+1}$$

تفاوت مورد انتظار در نسبت زوج‌های ناسازگار را با

$\mu = \pi_{01} - \pi_{10}$ و واریانس آن را با σ^2 نشان می‌دهیم و داریم [۸]:

$$\mu = \frac{\psi-1}{\psi+1} \quad \sigma^2 = \frac{4\psi}{(\psi+1)^2}$$

لذا با جای‌گذاری μ به جای δ و σ^2 به جای σ_δ^2 در معادله (۱) می‌توان تعداد زوج‌های ناسازگار را به دست آورد که با بازنویسی آن داریم:

$$n_d = \frac{4\psi \left(Z_{1-\frac{\alpha}{2}} + Z_{1-\beta} \right)^2}{(\psi-1)^2} + \frac{Z_{1-\frac{\alpha}{2}}^2}{2}$$

۲-۳) داده‌های زوجی چندحالتی اسمی. در این حالت نظیر حالت قبل، نسبت زوج‌های ناسازگار در نظر گرفته می‌شود و با در نظر گرفتن این مطلب که هر مورد و یا شاهد فقط در یکی از K سطح ممکن از مواجهه قرار می‌گیرد، نمادگذاری طبق جدول ۲ خواهد بود.

جدول ۲. نمادگذاری نتایج یک مطالعه مورد-شاهدی زوجی با پاسخ چندحالتی

شاهد مورد	با عامل ۱ مواجهه دارد	با عامل ۲ مواجهه دارد	...	با عامل k مواجهه دارد
با عامل ۱ مواجهه دارد	π_{11}	π_{12}	...	π_{1k}
با عامل ۲ مواجهه دارد	π_{21}	π_{22}	...	π_{2k}
⋮	⋮	⋮	⋮	⋮
با عامل k مواجهه دارد	π_{k1}	π_{k2}	...	π_{kk}

رسیدن به دو سطح برای پاسخ و سپس استفاده از روش بند قبل قابل اجرا می‌باشد. در این صورت برای هر ادغام یک حجم نمونه و در نهایت n_1 و n_2 و ... و n_K حجم نمونه خواهیم داشت که n بهینه از بین آن‌ها انتخاب خواهد شد (n بهینه می‌تواند بزرگترین n به دست آمده و یا n باشد که توان

نسبت بخت مورد به شاهد باشد، آنگاه برآورد تعداد زوج‌های ناسازگار برای نشان دادن این بخت در سطح

معنی‌داری دوطرفه α و توان $1-\beta$ از رابطه زیر به دست می‌آید که توسط کونت [۱۰] ارائه شده است: (۵)

$$n_d = \frac{\left(Z_{1-\frac{\alpha}{2}} (\psi+1) + Z_{1-\beta} \sqrt{(\psi+1)^2 - (\psi-1)^2 \pi_d} \right)^2}{(\psi-1)^2}$$

برای محاسبه تعداد کل زوج‌ها (که شامل زوج‌های سازگار و ناسازگار می‌باشد) بایستی n_d را بر $\pi_d = \pi_{01} + \pi_{10}$ (که احتمال مورد انتظار مشاهده یک زوج ناسازگار و یا ناجور

است) تقسیم نمود. بنابراین $n = \frac{n_d}{\pi_d}$ در عمل به جای n از برآورد آن‌ها یعنی P هایی استفاده می‌شود که از مطالعه مقدماتی و یا مطالعات مشابه به دست آمده‌اند.

یک روش دیگر براساس مشخص بودن ψ قرار دارد. زیرا با معلوم بودن ψ می‌توان نسبت‌های مورد انتظار زوج‌های

نسبت زوج‌های سازگار عبارت است از π_{11} و π_{22} و ... و π_{KK} . آزمون برابری دو K (کاپای) وابسته در این حالت مناسب است و تعیین حجم نمونه می‌تواند براساس آن انجام شود، اما تاکنون مورد بررسی قرار نگرفته و توسط نویسندگان مقاله تحت بررسی است. با این وجود یک روش ساده و در عین حال عملی با ادغام طبقات در یک دیگر تا

کلیه آزمون‌ها با در نظر گرفتن آن از یک حداقل مورد انتظار کم‌تر نباشد).

۴) نسبت شاهد-مورد بزرگتر از یک.

ممکن است یک مطالعه اپیدمیولوژی طوری طراحی شده باشد که در آن لازم باشد که تعداد افراد شاهده‌ی که با یک مورد جور می‌شوند، بیش از یک باشد. این مسأله هنگامی رخ می‌دهد که تعداد موارد زیاد نبوده، اما در عوض تعداد شاهدها به نسبت زیاد و سهل‌الوصول باشند. در چنین طرح مطالعه‌ای می‌توان با وجود نادر بودن موارد، توان مطالعه را افزایش داد [۱۷]. فرض کنید یک مورد با c شاهد جور شده باشد، لذا هر واحد جور شده شامل $1+c$ نفر خواهد بود. برای محاسبه تعداد واحدهای لازم، ابتدا می‌توان توسط معادلات ذکر شده در قسمت‌های قبل، تعداد زوج‌های حالت یک شاهد به یک

مورد یعنی n_E را محاسبه و سپس با تقسیم آن بر $1+c$ تعداد کل واحدها را به دست آورد. بنابراین

$$n_u = \frac{n_E(1+c)}{2c}$$

این صورت تعداد کل موارد عبارت است از n_u و تعداد کل شاهدها عبارت است از $n_c = cn_u$. بنابراین کل افرادی که در مطالعه وارد می‌شوند برابر است با

$$n_T = n_c + n_u$$

با این روش توان آزمون در سطح توان آزمون طرح یک به یک باقی خواهد ماند، با این تفاوت که تعداد موارد لازم با افزایش تعداد شاهدها کاهش یافته است.

نتایج براساس مثال‌های کاربردی

۱) خاتمی و همکاران [۱۸] برای بررسی نتایج درمانی چاقی مرضی توسط روش جراحی گاستروپلاستی عمودی (VGB) مطالعه‌ای را طراحی کردند که در آن ۳۰ بیمار تحت عمل جراحی قرار گرفتند. میزان موفقیت عمل با کاهش وزن بیماران پس از طی یک دوره یک‌ماهه سنجیده شد. میانگین و انحراف معیار کاهش وزن بیماران پس از درمان به ترتیب ۱۱/۰۳ و ۲/۴۸ کیلوگرم بوده است. هرگاه محقق دیگری درصدد تکرار این مطالعه بوده و بر این باور باشد که کاهش

وزن با روش مذکور به‌طور متوسط حداقل ۳ کیلوگرم خواهد بود، آن‌گاه برای اثبات این ادعا در سطح معنی‌داری ۵٪ و توان ۹۰٪ چه تعداد بیمار را باید مورد بررسی قرار دهد؟

با استفاده از معادله (۱) و جای‌گذاری پارامترهای $\alpha = 0.05$, $\beta = 0.1$, $\delta = 3$ و $\sigma_\delta = 2.48 \approx 2.5$ در آن نتیجه می‌شود $n = 11$ بیمار.

۲) محمد و همکاران [۱۹] با ارائه یک مثال به آزمون مقایسه فشارخون سیستول، قبل و پس از تجویز یک داروی خاص اشاره نموده‌اند، چنان‌چه محقق درصدد باشد تا با استفاده از اطلاعات مذکور آزمون مشابهی در سطح معنی‌داری دوطرفه ۰/۰۵ و توان ۰/۹۵ انجام دهد، باید چه تعداد نمونه را بررسی کند؟

به کمک جدول داده‌های ارائه شده در این منبع، برآورد P_1 عبارت است از مقدار ۰/۸۳ (یعنی تعداد حالت‌های مجزایی که $d_i + d_j > 0$ تقسیم بر تعداد کل حالت‌های مجزا اعم از بزرگ‌تر یا کم‌تر بودن این مجموع از صفر) در این صورت با در نظر گرفتن $\alpha = 0.05$, $\beta = 0.05$ و به‌کارگیری معادله (۲) به دست می‌آوردیم: $n \approx 40$ نفر.

۳) در یک کارآزمایی بالینی کاپچیک و همکارانش [۲۰] از ترکیب الکتروکانویولسیو-کلوزاپین برای درمان ۳۶ بیمار روانی استفاده نمودند و پس از پایان دوره درمان، بهبودی کامل را در ۲۴ نفر، خوب در ۳ نفر، نسبی در ۱ نفر، خفیف در ۴ نفر و عدم درمان را در ۴ نفر مشاهده نمودند. هرگاه محقق درصدد باشد تا مطالعه مذکور را در سطح معنی‌داری ۱٪ و توان ۹۹٪ تکرار کند، چه تعداد بیمار را بایستی مورد آزمون قرار دهد؟

از آن‌جا که این مطالعه به صورت قبل و بعد انجام شده است و همه نمونه‌ها در شروع مطالعه در شدیدترین درجه بیماری بوده‌اند با انتساب مقادیر ۰، ۱، ۲، ۳ و ۴ به شدت‌های بیماری از بیماری کامل تا سلامت کامل، مقادیر d_i ها که نشان‌دهنده شدت درمان هستند نیز چنین خواهند بود و π_i های مربوط به آن‌ها به ترتیب عبارتند از: ۰/۱۱، ۰/۱۱،

۰/۰۳، ۰/۰۸ و ۰/۶۷؛ بنابراین با به‌کارگیری معادله (۳) داریم:

$$\sigma^2 = 2.12 \text{ و } \mu \approx 3.09$$

از قرار دادن μ و σ^2 به ترتیب به جای δ و σ_δ^2 در معادله (۱) تعداد زوج‌های ناچور عبارت است از $n_d = 7$ و

لذا طبق رابطه (۴) تعداد کل زوج‌ها عبارت خواهد بود از:

$$n = \frac{kn_d}{k-1} = \frac{5}{4} \times 7 = 8.79 \approx 9$$

یعنی ۹ بیمار بایستی در این مطالعه قبل و بعد از درمان مورد آزمون قرار بگیرند.

(۴) در یک مطالعه مورد شاهدهی که توسط براون و همکارانش [۲۱] انجام شد، هدف بررسی اثر خودآزمایی سرطان بیضه بر روی میزان ابتلا به این سرطان بوده است. موارد کسانی بودند که مبتلا به سرطان بیضه و شاهدها به این بیماری مبتلا نبوده و با شاهدها جور شده بودند. جدول ۳ وضعیت موارد و شاهدها را با توجه به انجام خودآزمایی سرطان بیضه نشان می‌دهد:

جدول ۳. وضعیت موارد و شاهدها را با توجه به انجام خودآزمایی

سرطان بیضه

مورد	شاهد	انجام آزمون سرطان بیضه	عدم انجام آزمون سرطان بیضه
انجام آزمون سرطان بیضه	۴	۱۱	
عدم انجام آزمون سرطان بیضه	۳		۲۴۱

هرگاه محقق درصدد باشد که این مطالعه را به صورت مشابهی تکرار کند، برای خطای ۵ درصد و توان ۹۰ درصد چه تعداد زوج لازم دارد؟

از جدول ۳ داریم $\psi = \frac{11}{3}$ و برای $\alpha = 0.05$ و $1 - \beta = 0.9$ از فرمول (۵) به‌دست می‌آوریم:

$$n_d = 31.7 \approx 32$$

که از تقسیم این تعداد بر π_d داریم $n \approx 593$ زوج، یعنی در کل ۱۱۸۶ نمونه لازم است.

هرگاه به جای رابطه (۵)، حجم نمونه را از روابط (۶) و (۷) محاسبه کنیم، خواهیم داشت:

$$n_d = 23.44 \approx 24$$

و $n = 445$

اگر به‌دلایلی دسترسی به ۵۹۳ فرد مبتلا به سرطان بیضه مقدور نباشد و محقق درصدد باشد که با هر فرد بیمار ۱۰ نفر شاهد را جور کند، آن‌گاه با همان مفروضات حجم نمونه چقدر خواهد بود؟

در این حالت طبق روابط ذکر شده در بند ۲-۴ برای $c = 10$ ، $n_u = 327$ و $n_c = 3270$ لذا $n_t = 3270 + 327 = 3597$ نفر خواهد بود.

بحث و نتیجه‌گیری

در این مقاله به تفصیل به نحوه تعیین حجم نمونه در مطالعات زوجی و روابط آن‌ها در حالات گوناگونی که بر اساس ماهیت پاسخ مورد بررسی پدید می‌آیند پرداختیم. هم‌چنین سعی نمودیم تا با ارائه مثال‌های کاربردی نحوه استفاده از این روابط را تشریح کنیم. یک نکته اساسی در این بحث دخیل کردن هم‌بستگی بین زوج‌ها در محاسبه حجم نمونه‌ها می‌باشد، زیرا همان‌طور که استفاده از آزمون‌هایی که بر مبنای نمونه‌های مستقل انجام می‌شوند به دلیل چشم‌پوشی از این هم‌بستگی‌ها پایستار (Conservative) بوده و می‌توانند منجر به نتایج گمراه‌کننده‌ای شوند، استفاده از روابط تعیین حجم نمونه برای گروه‌های مستقل نیز می‌تواند به طور کاذب باعث افزایش یا کاهش تعداد نمونه‌ها شود که در حالت اول باعث اتلاف هزینه و زمان و گاهی نیز در معرض خطر قرار دادن تعداد بیش‌تری از بیماران شود و در حالت دوم باعث کاهش صحت و اعتبار نتایج به دست آمده شود. برای روشن‌تر شدن این مطالب فرض کنید در یک مطالعه جور شده هدف مقایسه میانگین دو گروه برای نشان دادن اختلافی به اندازه $\delta = 2$ در سطح معنی‌داری دوطرفه $\alpha = 0.05$ و توان $1 - \beta = 0.95$ باشد. در این صورت اگر واریانس مشترک دو گروه به طور مثال $\sigma^2 = 4$ باشد، آن‌گاه جدول ۴ و نمودار ۱ تغییرات حجم نمونه

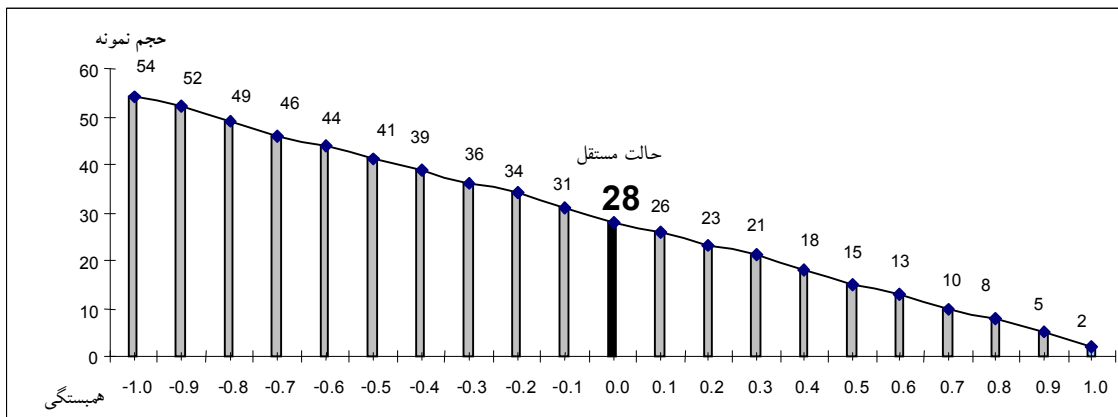
لازم (با استفاده از نتایج قسمت ۱-۱) را برحسب تغییرات P یا ضریب هم‌بستگی بین دو گروه نشان می‌دهند.

جدول ۴. تغییرات حجم نمونه لازم برحسب تغییرات P یا ضریب هم‌بستگی بین دو گروه

مقدار هم‌بستگی	۱	۰/۹	۰/۸	۰/۷	۰/۶	۰/۵	۰/۴	۰/۳	۰/۲	۰/۱	۰	-۰/۱	-۰/۲	-۰/۳	-۰/۴	-۰/۵	-۰/۶	-۰/۷	-۰/۸	-۰/۹	-۱
تعداد نمونه*	۲	۵	۸	۱۰	۱۳	۱۵	۱۸	۲۱	۲۳	۲۶	۲۸	۳۱	۳۴	۳۶	۳۹	۴۱	۴۴	۴۶	۴۹	۵۲	۵۴

*به عدد بالا گرد شده اند

نمودار ۱. تغییرات حجم نمونه لازم برحسب تغییرات P یا ضریب هم‌بستگی بین دو گروه



روش‌های ارائه شده در این مقاله را می‌توان با در نظر گرفتن پاره‌ای ملاحظات برای مواقعی که با مطالعات تقاطعی سرو کار داریم نیز به کار گرفت. خواننده علاقه‌مند می‌تواند برای آشنایی بیشتر با جزئیات لازم به منابع [۲۳، ۲۴] مراجعه نماید.

با بررسی این جدول و نمودار مشاهده می‌شود که با افزایش هم‌بستگی مثبت بین دو گروه حجم نمونه نسبت به حالت مستقل ($P=0$) رو به کاهش می‌گذارد و برعکس با افزایش هم‌بستگی منفی، رو به افزایش می‌گذارد. بدون این‌که نیازی به شبیه‌سازی باشد مشاهده می‌شود که بدون دخیل کردن هم‌بستگی، چنان‌چه این هم‌بستگی مثبت باشد، تعداد نمونه برحسب استقلال دو گروه بیش از آن‌چه لازم است و برعکس اگر این هم‌بستگی منفی باشد، تعداد نمونه کم‌تر از تعداد لازم خواهد بود که منجر به کاهش توان آزمون خواهد شد. البته همان‌طور که در بند ۱-۱ ذکر شد اعتبار این رابطه تا آن‌جاست که P در بازه عددی $(-0/75, 0/75)$ قرار داشته باشد؛ یعنی در این حالت خاص تعداد نمونه دیگر نمی‌تواند کم‌تر از ۹ و بیش‌تر از ۴۸ باشد.

در حالتی که با پاسخ‌های رتبه‌ای سروکار داریم آدکاک [۲۲] روشی را پیشنهاد نموده که بر مبنای روش‌های بیزی عمل می‌کند. با این وجود از آن‌جایی که روش‌های بیزی به دلیل پیچیدگی‌هایشان در تعیین تابع پیشین، هنوز جایگاه مستحکمی در مطالعات پزشکی (به‌خصوص در بین متخصصین غیرآماري) نیافته‌اند، لذا استفاده از آن‌ها به سادگی میسر نمی‌باشد.

منابع

- [1] Rosner B. Fundamentals of Biostatistics. 3rd ed. Massachusetts: PWS-Kent Publishing Co, 1990.
- [2] Fliess JL. The design and analysis of clinical experiments. New York: John Wiley & Sons, 1999.
- [3] Connet JE, Smith JA, McHugh RB. Sample size and power for pair-matched case-control studies. *Statistics in Medicine*, 1987; 6:53-9.
- [4] Bland JM, Altman DG. Matching. *BMJ*, 1994; 309(6962):1128.
- [5] Senn S. The design and analysis of cross-over trials. Wiley, 1992.
- [6] Schlesselman JJ. Case-control studies. New York: Oxford University Press, 1982.
- [7] Campbell MJ, Machin D. Medical statistics: A commonsense approach. 2nd ed. Chichester: John Wiley & Sons Ltd; 1992.
- [8] Lachenbruch PA. On the sample size for studies based upon mcnemar's test. *Statistics in Medicine*, 1992; 11:1521-5.
- [9] Machin D, Campbell M, Fayers P, Pinol A. Sample size tables for clinical studies, 2nd Ed. London, Edinburgh, Malden and Carlton: Blackwell Science, 1997.
- [10] Connor RJ. Sample size for testing differences in proportions for the paired-sample design. *Biometrics*, 1987; 43(1):207-11.
- [11] Guenther WC. Sample size formulas for normal theory t tests. *American Statistician*, 1981; 35(4):243-4.
- [12] Kupper LL, Hafner KB. How appropriate are popular sample size formulas? *American Statistician*, 1989; 43:101-5.

- [۱۹] محمد کاظم، ملک‌افضلی حسین، نهایتیان وارثکس. روش‌های آماری و شاخص‌های بهداشتی. چاپ نهم. تهران: ابلاغ، ۱۳۷۷، ص ۱۲۱-۱۱۸.
- [20] Kupchick M, Spivak B, Mester R, Reznik I, Gonen N, Weizman A, et al. Combined electroconvulsive-clozapine therapy. *Clinical Neuropharmacology*, 2000; 23(1):14-6.
- [21] Brown LM, Pottern LM, Hoover RN. Testicular cancer in young men: the search for causes of the epidemic increase in the United States. *J Epidemiol Commun Health*, 1987; 41:349-54.
- [22] Adcock CJ. The bayesian approach to determination of sample size-some comments on the paper by Joseph Wolfson and du Berger. *Statistician*, 1995; 44:155-61.
- [۲۳] سلیمانی بهرام، فقیه زاده سقراط، کاظم نژاد انوشیروان، بابایی غلامرضا. تاثیر انتقالی بر حجم نمونه و توان آزمون در مطالعات متقاطع ۲×۲. پژوهش در علوم پزشکی، ۱۳۸۲، دوره ۸، خرداد و تیر، شماره ۲، ص ۱۶-۱۱.
- [24] Lachin JM. Introduction to sample size determination and power analysis for clinical trials. *Control Clin Trials*, 1981; 2:93-113.
- [13] Frison L, Pocock SJ. Repeated measures in clinical trials: analysis using mean summary statistics and its implications for design. *Stat Med*, 1992; 11(13):1685-704.
- [14] Noether GE. Sample size determination for some common nonparametric tests, *J Am Statist Assoc*, 1987; 82(398):643-647.
- [15] Parker RA, Bergman DJ. Sample size for individually matched case-control studies. *Biometrics*, 1986; 42:919-26.
- [16] Royston P. Exact conditional and unconditional sample size for pair-matched studies with binary outcome: a practical guide. *Stat Med*, 1993; 12(7):699-712.
- [17] Miettinen OS. Individual matching with multiple controls in the case of all-or-none responses. *Biometrics*, 1969; 25(2):339-55.
- [۱۸] خاتمی سیدمسعود، مهرروز شعبان، پناهی فرزاد، اسعدی مهدی. بررسی نتایج درمانی و عوارض عمل جراحی گاستروپلاستی عمودی (VGB) در چاقی مرضی. کومش، مجله علمی پژوهشی دانشگاه سمنان، ۱۳۸۳، دوره ۵، بهار و تابستان، شماره ۳-۴، ص ۱۶۲-۱۵۷.