بررسی تأثیر اندازه‌ی نمونه در مدل‌های چندسطحی با رویکرد زیرنمونه‌گیری

امید اخگری و موسی گل‌علی‌زاده[1]

گروه آمار دانشگاه تربیت مدرس

تاریخ دریافت: ۹۲/۴/۱ تاریخ پذیرش: ۹۲/۸/۳۰

چکیده: موضوعات تحقیقاتی بی‌شماری در حوزه علوم اجتماعی، پزشکی، کشاورزی و غیره وجود دارند که شامل داده‌هایی با ساختار همبستگی درون‌گروهی است. واضح است برای چنین داده‌هایی مدل‌های رگرسیون خطی معمولی به‌دلیل عدم لحاظ این همبستگی ذاتی از کارایی قابل قبولی برخوردار نیستند. در این حالت مدل‌های مناسب که قابلیت لحاظ نمودن همبستگی درون‌گروهی و استقلال بین‌گروهی را دارند به مدل‌های چندسطحی معروف هستند. یکی از موضوعات اساسی مطالعه این‌گونه مدل‌ها یافتن روشی برای تعیین اندازه نمونه مناسب در سطوح مختلف آن است. اما به دلیل پیچیدگی‌های موجود به‌جای جستجوی روش تعیین اندازه نمونه با استفاده از رویکردهای خاص به بررسی تأثیر اندازه نمونه بر عملکرد برآوردهای مدل می‌پردازند. در این مقاله با رویکرد روش زیر نمونه‌گیری تأثیر اندازه‌های مختلفی از اندازه نمونه از سطح دوم روی برآورد اثرهای ثابت و تصادفی مورد مطالعه قرار گرفت. به علاوه به دلیل ارتباط تنگاتنگ موضوع تعیین اندازه نمونه و توان آزمون آماری مربوط به پارامترهای مورد مطالعه و همچنین عوامل دیگری از قبیل اندازه‌ی اثر و سطح معنی‌داری، با استفاده از شبیه‌سازی به ارزیابی طرح اندازه‌های مختلف نمونه‌ای پرداخته می‌شود. نتایج حاصل حاکی از این است که افزایش اندازه نمونه در سطح دوم مدل دوسطحی توان آزمون مربوط به برآورد پارامتری ثابت و تصادفی را افزایش می‌دهد.

واژه‌های کلیدی: مدل‌های چندسطحی، اندازه‌ی نمونه‌ی بهینه، اثرهای ثابت و تصادفی، طرح زیرنمونه‌گیری؛ توان آزمون آماری.

رده‌بندی ریاضی: ۶۲J۱۲، ۶۲K۰۵.

۱- آدرس الکترونیکی نویسنده مسئول مقاله: موسی گل‌علی‌زاده golalizadeh@modares.ac.ir

۱- مقدمه

مثال‌های بی‌شماری از حوزه‌های مربوط به علوم پزشکی، اجتماعی، کشاورزی و سایر علوم وجود دارند که داده‌های مورد مطالعه در آن‌ها به‌صورت درون‌گروهی همبسته هستند؛ در حالی که گروه‌های مختلف مستقل از هم می‌باشند. گرچه تحلیل آماری چنین داده‌هایی با روش‌های رگرسیون خطی ساده قابل انجام است، اما با توجه به این‌که یکی از فرض‌های اساسی این‌گونه مدل‌های رگرسیونی استقلال آماری بین مشاهدات است، نقض این فرض سبب می‌شود که خطای استاندارد پارامترهای موجود در مدل به‌طور بالقوه کم برآورد شوند [۱]. به منظور لحاظ نمودن این ویژگی، از مدل‌هایی با عنوان مدل‌های چندسطحی استفاده می‌شود که به‌طور مناسبی همبستگی درون‌گروهی و استقلال بین‌گروهی را در نظر می‌گیرد [۲]. به‌دلیل کاربرد وسیع مدل‌های چندسطحی در علوم مختلف مبانی تئوریک و محاسباتی گوناگونی از این مدل-ها در طی سال‌های اخیر مورد مطالعه قرارگرفته که نتایج آن‌ها منجر به نرم‌افزارهای آماری بسیار قوی مختص مدل‌های چندسطحی مثل MLwiN شده است [۳].

محققین در سایر علوم برای جمع‌آوری داده‌های خود نیازمند توجه به این نکته هستند که برای دستیابی به نتیجه مطلوب باید دقت مطالعه آزمایشی تا حد ممکن افزایش یابد و یکی از عوامل تأثیرگذار برای این مهم افزایش اندازه‌ی نمونه‌ی مورد مطالعه است. اما گاهی اوقات به علت محدودیت‌هایی که به خصوص در زمینه هزینه، زمان و غیره موجود است نمی‌توان از نمونه‌ی زیادی استفاده نمود. این موضوع در مدل‌های چندسطحی نیز به شکل‌های مختلف مطرح شده است [۱ و ۳]. نکته قابل تأمل در خصوص داده‌های چندسطحی این است که نحوه انتخاب و اندازه نمونه در سطح اول و دوم منجر به نتایج متفاوتی خواهد شد. این واقعیت مسئله تعیین اندازه‌ی نمونه‌ی بهینه را در مدل‌های چندسطحی جذاب‌تر نموده است. برای ارائه تصویری از فعالیت‌های صورت گرفته در زمینه تعیین اندازه‌ی نمونه‌ی بهینه در مدل‌های چندسطحی خلاصه‌ای از آن‌ها در ذیل آمده است.

بعد از معرفی مدل‌های چندسطحی در سال ۱۹۸۷، محققین مختلفی از حوزه‌های متفاوت علوم در زمینه‌ی تعیین اندازه‌ی نمونه در این مدل‌ها فعالیت کردند [۴]. از جمله آن‌ها می‌توان به گلداستین و سیلور اشاره نمود [۵]. به‌علاوه، بر اساس انحراف معیار پارامترهای یک مدل دوسطحی و نوع خاصی از معادله هزینه، اسنایدر و بوسکر نیز فرمول‌هایی برای تعیین اندازه‌ی نمونه و ارتباط آن با توان آزمون‌های مربوط به پارامترهای مدل ارائه کردند [۶ و ۷]. همچنین با استفاده از الگوریتم خودگردان سازی[۱]، افشار توس در [۸] به ارزیابی تأثیر اندازه‌ی نمونه در سطح دوم، هنگامی که تعداد واحدهای سطح اول کم باشد، پرداخت. برای پاسخ به بخشی از

1- Bootstrapping

سؤالات مرسوم در زمینه تعیین اندازه‌ی نمونه در مدل‌های دوسطحی کوهن راهکارهای بسیار مناسب را ارائه نموده است [۱]. اخیراً براون و همکاران در [۳]، به کمک شبیه‌سازی ترکیبات نمونه‌ای متفاوت، اما بدون لحاظ نمودن عامل هزینه، موضوع تعیین اندازه‌ی نمونه را برای مدل‌های مختلف چندسطحی مورد مطالعه قرارداد و نرم‌افزاری را نیز در این زمینه تولید کردند.

رویکرد اکثر فعالیت‌های صورت گرفته در زمینه تعیین اندازه‌ی نمونه بهینه در مدل‌های چند-سطحی برای تأثیر اندازه نمونه‌های مختلف و برآورد پارامترهای مدل با استفاده از شبیه‌سازی بوده است. مقاله حاضر نیز با انجام شبیه‌سازی به‌صورت زیرنمونه‌گیری در یک مدل دوسطحی تأثیر تعداد واحدهای سطح دوم را روی برآورد اثرهای ثابت و تصادفی مورد ارزیابی قرار می-دهد. استفاده از رویکرد زیرنمونه‌گیری وجه تمایز این مقاله با تحقیق‌های قبلی در این زمینه است. ساختار مقاله حاضر به‌صورت زیر است. ابتدا در بخش دوم خلاصه‌ای گذرا از مدل‌های چندسطحی ارائه می‌شود. با یادآوری ارتباط بین توان آماری و اندازه‌ی نمونه مناسب مشکلات پیش‌روی تعیین اندازه‌ی نمونه‌ی بهینه در مدل‌های چندسطحی در بخش سوم مورد مطالعه قرار می‌گیرد. با استفاده از یک مثال کاربردی در بخش چهارم تأثیر تعداد گروه در دقت برآوردها و توان آماری مربوطه با استفاده از روش زیرنمونه‌گیری مورد بررسی قرار می‌گیرد. مقاله با بحث و نتیجه‌گیری کلی خاتمه می‌یابد.

۲- مروری گذرا بر مدل‌های چندسطحی

گرچه مدل‌های چندسطحی به‌صورت غیر رسمی در سالیان بسیار دور مورد استفاده قرار می-گرفت اما به‌طور رسمی و برای اولین بار این مدل توسط گلداستین در سال ۱۹۸۹ برای مدل-بندی داده‌های خطی معرفی شد [۵]. پس از آن کاربردهای وسیع چنین مدل‌هایی توسط محققین علوم مختلف شامل پزشکی، علوم اجتماعی و کشاورزی به گسترش و بسط آن کمک شایانی کرد. امروزه تلفیق این مدل با مدل‌های گوناگون آماری دیگر منجر به دستاوردهای جدیدی در مدل‌بندی پدیده‌های مختلف شده است. می‌توان تصور کرد که بنا به نیاز علوم متفاوت نمادگذاری‌های متنوعی برای تحلیل مدل‌های چندسطحی معرفی شد اما در مقاله‌ی حاضر از نمادگذاری گلداستین استفاده می‌شود [۲]. برای ادامه بحث تعدادی از نمادگذاری‌های مرسوم مدل‌های چندسطحی در زیر می‌آید:

فرض کنیم N فرد درون J گروه، گروه‌بندی شده باشند، طوری که n_j تعداد افراد گروه jام است. به‌علاوه فرض کنیم که بخواهیم برای J گروه، متغیر (پیوسته) پاسخ y_j را روی p دسته از متغیرهای پیوسته که در ماتریس X_j ذخیره شده‌اند، رگرسیون نماییم. از این‌رو، برای jامین گروه می‌نویسیم:

$$Y_j = X_j \beta_j + r_j, \qquad j = 1,\ldots,J \qquad (1)$$

که Y_j برداری با n_j بعد، X_j ماتریسی با ابعاد $n_j \times p$، β_j بردار پارامتر p بعدی و r_j خطای اندازه‌گیری و یک متغیر تصادفی است به‌طوری که

$$r_j \sim N\left(0_{n_j}, \sigma^2 I_{n_j}\right), \qquad (2)$$

که I_p ماتریس همانی p بعدی و برداری n_j بعدی از صفرها است. در ادبیات مدل‌های چندسطحی مدل (۱) همراه با فرض (۲) به مدل‌های سطح اول اشاره دارد. مدل‌های دوسطحی با در نظر گرفتن β_j در مدل (۱) به‌عنوان یک متغیر تصادفی شکل خواهد گرفت. یک حالت خاص مدل دوسطحی این است که β_j به‌صورت زیر مدل‌بندی شود:

$$\beta_j = W_j \gamma + u_j, \qquad (3)$$

که W_j ماتریس متغیرهای تبیینی در سطح گروه jام، γ برداری از ضرایب ثابت و u_j بردار خطا و دارای توزیع نرمال با میانگین صفر و واریانس T است؛ یعنی $u_j \sim N(0,T)$. بنابراین می‌توان β_j را به‌عنوان متغیری تصادفی دارای توزیع نرمال با میانگین $W_j \gamma$ و واریانس T در نظر گرفت.

ترکیب فرمول‌های سطح اول و دوم مدل دوسطحی روابط (۱) و (۳) عبارت است از:

$$Y_j = X_j W_j \gamma + X_j u_j + r_j, \quad i = 1,\ldots,J. \qquad (4)$$

بنابراین Y_j دارای توزیع نرمال با میانگین $X_j W_j \gamma$ و ماتریس واریانس $X_j T X_j' + \sigma^2 I_{n_i}$ است. بنابه رابطه‌ی (۴) واضح است که Y_j بر اساس هر دو اثرات ثابت γ و اثرهای تصادفی u_j و r_j مدل‌بندی شده است. به‌همین دلیل است که این مدل‌ها را به نام مدل‌های آمیخته نیز می‌شناسند [۲].

به‌طور کلی مدل‌های چندسطحی در سه کلاس متمایز گنجانده می‌شوند: ساده‌ترین مدل، که مدل عرض از مبدأ تصادفی بدون حضور متغیر مستقل (مدل مؤلفه واریانس) است. مدل دوم مدلی با عرض از مبدأ تصادفی و با حضور متغیرهای مستقل بوده و مدل سوم که نسبت به دو مدل دیگر پیچیده‌تر است مدلی با عرض از مبدأ و شیب تصادفی خواهد بود.

واضح است که ترکیبی از گونه‌های متفاوت چنین مدل‌ها وقتی‌که تعداد زیادی متغیرهای تبیینی در مدل حاضر باشد نیز قابل ارائه خواهد بود. به‌عنوان مثال رابطه زیر یک مدل عرض از مبدأ و شیب تصادفی دوسطحی برای متغیر پاسخ y با متغیر تبیینی x در سطح اول و متغیر تبیینی w (برای توضیح متغیر پاسخ سطح دوم یا ضرایب) در سطح دوم است:

$$y_{ij}=\beta_{0j}+\beta_{1j}x_{ij}+\epsilon_{ij},\quad i=1,\ldots,n \quad j=1,\ldots,J, \text{ سطح اول}$$

$$\text{سطح دوم}\quad \begin{aligned}\beta_{0j}&=\gamma_{00}+\gamma_{01}w_j+u_{0j},\\ \beta_{0j}&=\gamma_{10}+\gamma_{11}w_j+u_{1j}.\end{aligned} \qquad (۵)$$

نحوه‌ی ارائه مدل در روابط (۵) نه تنها همه‌ی متغیرهای وابسته را به ما نشان می‌دهد بلکه به روشنی نشان‌دهنده‌ی ماهیت چندسطحی مدل است. به‌علاوه در معادلات (۵) تفکیک سطوح اول و دوم از هم به وضوح مشخص شده است؛ به این ترتیب که سطح اول نشان‌دهنده‌ی یک رگرسیون خطی معمولی است و سطح دوم نحوه‌ی ارتباط پارامترهای سطح یک با متغیرهای سطح دوم را نشان می‌دهد.

مطالب بی‌شماری در خصوص استنباط‌های آماری و موضوعات دیگر راجع به مدل‌های فوق وجود دارند که بیان آن‌ها از حوصله‌ی بحث این مقاله خارج است. به‌عنوان مثال برای برآورد پارامترهای مدل می‌توان از روش‌های ماکسیمم درستنمایی، حداقل مربعات تعمیم‌یافته تکراری و حداقل مربعات تعمیم‌یافته تکراری مقید استفاده نمود [۹]. خواننده علاقه‌مند می‌تواند به اسنایدر و بوسکر [۷]، گلمن و هیل [۹] و گلداستین [۲] مراجعه نماید.

۳- تعیین اندازه‌ی نمونه در مدل‌های چندسطحی

همان‌گونه که کوهن در [۱] اشاره کرد، محاسبه‌ی اندازه‌ی نمونه فرایندی نسبتاً پیچیده است؛ چرا که محاسبه‌ی اندازه‌ی نمونه بر مبنای فرض‌هایی است که بنا به شرایط خاص قابل تغییرند. این شرایط می‌تواند گسستگی یا پیوستگی متغیر پاسخ، توان آزمون، دقت برآورد اثرهای ثابت و تصادفی باشد. با این حال یکی از پارامترهای تأثیرگذار در تعیین اندازه‌ی نمونه عامل هزینه است که معمولاً در اکثر بررسی‌های نمونه‌ای نادیده گرفته می‌شود. به عنوان مثال چون در مطالعات پزشکی اغلب به‌دست آوردن نمونه مورد نیاز مشکل و پرهزینه است محقق خود را محدود به اندازه‌ی نمونه تجربی می‌نماید (تعداد نمونه‌ای را که بر اساس بودجه معین در تحقیق به‌دست می‌آید، اندازه‌ی نمونه تجربی گویند). علی‌رغم این، اگر بتوان حداقل نمونه‌ی مورد نیاز برای دستیابی به دقت مورد نظر در برآورد پارامترهای مورد مطالعه را طوری یافت که توان آزمون مربوطه حداکثر شود می‌توان انتظار داشت که در هزینه و مشکلات مرتبط با آن صرفه‌جویی صورت خواهد گرفت [۱۰]. بیان این نکته ضروری است که در مقاله حاضر ما نیز عامل هزینه را نادیده گرفته‌ایم. اما به‌منظور نیل به هدف تعیین اندازه‌ی نمونه در مدل‌های چندسطحی، نیازمند مقدماتی ساده از آمار مقدماتی هستیم که در پی می‌آید.

مسئله آزمون فرض زیر را که در آن $\mu_1<\mu_0$، در نظر بگیریم:

$$\begin{aligned} H_0 &: \mu = \mu_0 \\ H_1 &: \mu = \mu_{10} \end{aligned} \tag{۶}$$

از آمار مقدماتی می‌دانیم که ناحیه‌ی رد این آزمون با شرط معلوم بودن σ^2، به‌صورت

$$\{x \mid \bar{x} < \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha}\}$$

است و داریم:

$$1-\beta = P\left(\bar{x} < \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha} \right). \tag{۷}$$

آنگاه به کمک تساوی (۷) می‌توان اندازه‌ی نمونه مورد نیاز (n) را به‌صورت زیر محاسبه نمود:

$$n = \frac{\left(z_{1-\beta} - z_{1-\alpha}\right)^2 \sigma^2}{\left(\mu_1 - \mu_0\right)^2}. \tag{۸}$$

این تساوی رابطه تنگاتنگ اندازه‌ی نمونه، سطح اطمینان، اندازه اثر و توان آزمون $(1-\beta)$ را نشان می‌دهد.

اساساً، در اکثر محاسبات مربوط به توان و اندازه‌ی نمونه چهار کمیت وابسته وجود دارد: اندازه-ی آزمون، توان آزمون، اندازه‌ی اثر (قدر مطلق تفاضل برآورد از مقدار واقعی آن) و انحراف استاندارد اندازه‌ی اثر (که تابعی از اندازه‌ی نمونه است). حال فرض کنیم بخواهیم فرض صفر $\gamma = 0$ را که یک ضریب رگرسیونی است آزمون کنیم و می‌دانیم که برای این ضریب یک برآورد $\hat{\gamma}$ که دارای توزیع تقریبی نرمال با انحراف استاندارد $s.e(\hat{\gamma})$ است، وجود دارد. اگر مقدار واقعی اندازه‌ی اثر γ بزرگ باشد و انحراف استاندارد آن کوچک باشد، آنگاه توان آزمون بزرگ خواهد بود و سطح معنی‌داری بالاتر به توان آزمون بزرگ‌تر سوق داده می‌شود. این واقعیت ناشی از رابطه‌ی تقریبی

$$\frac{\hat{\gamma}}{s.e(\hat{\gamma})} = z_{1-\beta} - z_{1-\alpha} \tag{۹}$$

است که با توجه به یک آزمون یک-طرفه به‌دست آمده است [۱۱]. در این رابطه α سطح معنی‌داری و β خطای نوع دوم را نشان می‌دهد. به‌علاوه z_q بیانگر صدک qام توزیع نرمال استاندارد است. بیان این نکته ضروری است که برای آزمون دوطرفه رابطه (۹) به فرمول تقریبی زیر تغییر می‌یابد:

$$\frac{\hat{\gamma}}{s.e(\hat{\gamma})} = z_{1-\beta} - z_{1-\alpha/2}. \qquad (10)$$

معمولاً در مدل‌های چندسطحی، اندازه‌ی نمونه برای هر سطح، به عنوان تعداد کل واحدهای مشاهده‌شده برای آن سطح تعریف می‌شود. آن‌گاه اندازه‌ی نمونه کل با ترکیب اندازه نمونه‌ها در تمامی سطوح به دست می‌آید. به‌همین دلیل، پرسش اصلی اندازه‌ی نمونه راجع به نمونه کل نبوده بلکه مربوط به تعداد نمونه در هر سطح خاص است [۱۲]. جهت روشن شدن مطلب یک مدل سه‌سطحی را که دانش‌آموزان درون کلاس‌ها و کلاس‌ها درون مدارس آشیانه شده‌اند، در نظر می‌گیریم. فرض کنیم تعداد ۳۳۰۰ دانش‌آموز در ۱۵۰ کلاس درس و در۶۰ مدرسه توزیع شده‌اند. بنابراین، به‌طور متوسط در هر کلاس درس ۲۲ دانش‌آموز و در هر مدرسه حداقل ۲ کلاس درس موجود است. یک محقق برای ارزیابی یک فرضیه علاقه‌مند است بداند چه تعداد دانش‌آموز از چند کلاس و از چند مدرسه انتخاب نماید؟

یکی از محدودیت‌های اساسی در تعیین اندازه‌ی نمونه در مدل‌های چندسطحی مربوط به اندازه‌ی نمونه آخرین (بالاترین) سطح است. به بیانی دقیق‌تر در مدل‌های با ساختار سلسله-مراتبی داشتن واحدهای تا حد ممکن زیاد در بالاترین سطح ضروری است [۸]. به دلیل ماهیت مدل‌های چندسطحی تعیین اندازه‌ی نمونه وابسته به اندازه آن در سطوح مختلف است. به عنوان مثال اگر بخواهیم تحت استانداردهای بین‌المللی و با کمک یک مدل دوسطحی، تعدادی داده‌های آموزشی جمع‌آوری نماییم، نیازمند تعیین اندازه‌ی نمونه بر اساس دو محدودیت زیر هستیم:

۱. تعداد مدرسه لازم در بررسی نمونه‌ای؛

۲. تعداد دانش‌آموزان مورد نیاز در هر مدرسه.

اضافه بر مواردی مثل حالت فوق تعیین اندازه‌ی نمونه به نوعی وابسته به هدف اصلی مدل‌بندی نیز می‌باشد [۶]. به‌علاوه در بعضی از مسائل تعیین اندازه‌ی نمونه، محقق علاقه‌مند به مطلوبیت برآورد پارامتر، قرار گرفتن برآورد در نواحی خاص و یا محدودیت‌های بیزی است. می‌توان تصور کرد لحاظ تمامی اهداف یک تحقیق در فرآیند تعیین اندازه‌ی نمونه بهینه در مدل‌های چندسطحی تا حدودی غیرممکن است. به همین خاطر در این مقاله اولاً بررسی پایایی اثرهای ثابت تحت اندازه‌های نمونه‌ای مختلف (اندازه‌ی نمونه‌ای کل و اندازه‌ی نمونه‌ای در سطح دوم) و ثانیاً بررسی پایایی مؤلفه‌های کواریانس به عنوان اهداف اصلی و تأثیرگذار در تعیین اندازه‌ی نمونه مدنظر قرار می‌گیرد.

۴- مطالعه یک مثال کاربردی

داده‌های مورد مطالعه در این مقاله از درگاه نرم‌افزار R با نام $bh1996$ از کتابخانه‌ی $Multilevel$ استخراج شده است. این داده‌ها که توسط بلیس و هالورسون در سال ۱۹۹۶ جمع‌آوری شد، شامل سه متغیر خلق و خوی کارفرما ($LEAD$)، کار مفید ($WEBING$) و ساعت کاری (HRS) است [۱۳]. قابل ذکر است که هر کدام از این متغیرها دارای مشتقات دیگری می‌باشند که از بین آن‌ها تنها از $G.HRS$ که میانگین ساعت کاری برای هر گروه است، استفاده شده است. بلیس و هالورسون این داده‌ها را بر اساس خلق و خوی کارفرما گروه‌بندی کرده‌اند. به این ترتیب که افراد تحت نظر هر کارفرما داده‌های سطح اول و کارفرماها داده‌های سطح دوم را تشکیل می‌دهند. از آنجا که تعداد کارفرماها ۹۹ است، واضح است که داده‌ها درون ۹۹ گروه، گروه‌بندی می‌شوند. قابل ذکر است که به‌منظور سهولت در تفسیر و شبیه‌سازی، تمام متغیرهای مورد مطالعه در این مقاله استاندارد شد. با توجه به اینکه داده‌ها دارای ساختار دو-سطحی است و با پیروی از بلیس و هالورسون، [۱۳]، برای $j=۱,\dots,۹۹$ و $i=۱,\dots,۷۳۸۲$ مدل مورد مطالعه به‌صورت زیر در نظر گرفته می‌شود:

سطح اول
$$WBEING_{ij} = \beta_{\circ j} + \beta_{۱j}\, HRS_{ij} + \beta_{۲j}\, LEAD_{ij} + \epsilon_{ij},$$

سطح دوم
$$\begin{aligned} \beta_{\circ j} &= \gamma_{\circ\circ} + \gamma_{\circ ۱}\, G.HRS_j + u_{\circ j}, \\ \beta_{۱j} &= \gamma_{۱\circ}, \\ \beta_{۲j} &= \gamma_{۲\circ}. \end{aligned}$$

با برازش این مدل در نرم‌افزار آماری R، نتایجی به‌صورت خلاصه‌شده در جدول ۱ به‌دست آمده است.

با توجه به این‌که تمامی p-مقدارها به جز p-مقدار عرض از مبدأ کمتر از پنج درصد است، می‌توان نتیجه گرفت که همه اثرهای ثابت جز عرض از مبدأ در سطح ۵٪ معنی‌دار هستند. به‌علاوه بنا به جدول ۱ مشاهده می‌شود که با کاهش ساعت کاری هر فرد (HRS) و افزایش متوسط ساعت کاری هر گروه ($G.HRS$)، کار مفید هر کارگر ($WEBING$) کاهش می‌یابد. به علاوه، بنا به این جدول خلق و خوی کارفرما ($LEAD$) نسبت به عوامل دیگر تأثیر مستقیم و زیادی روی کار مفید کارگران دارد.

ماتریس کواریانس برآورد شده برای مدل (۱۱) به‌صورت

$$\hat{\Gamma} = \begin{pmatrix} \hat{\sigma}^{۲}_{u\circ} & \hat{\sigma}_{\circ ۲} \\ \hat{\sigma}_{\circ ۲} & \hat{\sigma}^{۲}_{u۲} \end{pmatrix} = \begin{pmatrix} (\cdot/\cdot ۳۵)^{۲} & -\cdot/\cdot ۱۱ \\ -\cdot/\cdot ۱۱ & (\cdot/\cdot ۳۱)^{۲} \end{pmatrix} \tag{۱۲}$$

است که در آن $\hat{\sigma}^{۲}_{u\circ}$ برآورد واریانس خطاهای عرض از مبدأ در سطح دوم، $\hat{\sigma}^{۲}_{u۲}$ برآورد

واریانس خطاهای شیب $LEAD$ در سطح دوم و $\hat{\sigma}_{\circ ۲}$ برآورد کوواریانس بین خطاهای عرض از مبدأ و شیب $LEAD$ است. اضافه بر این برآورد واریانس خطای عرض از مبدأ عبارت است از:

$$\hat{\sigma}_e^{۲} = \cdot / ۸۸. \qquad (۱۳)$$

هدف این مقاله ارائه نحوه برازش مدل و ارزیابی کارایی آن نیست، بلکه هدف اصلی ما از ارائه برآوردهای فوق در نظر گرفتن سنگ محکی برای ارزیابی برآورد پارامترها به ازای اندازه‌ی نمونه های مختلف از اندازه‌ی کل نمونه در هر سطح اول و دوم است. به عبارتی دقیق‌تر، با تفسیر ترکیب‌های نمونه‌ای متفاوت از سطوح مختلف مدل دوسطحی (۱۱) نحوه تأثیر آن‌ها را در برآورد پارامترهای مدل مورد مطالعه قرار می‌دهیم. می‌توان این هدف را به طریق دیگر و با طرح سؤالی به‌صورت زیر مورد بررسی قرارداد: آیا تعداد نمونه‌های در نظر گرفته‌شده در برازش مدل (۱۱) کافی بوده است؟ روش‌های مناسبی برای پاسخ به سؤالی مثل این پرسش برای مدل‌های کلی دوسطحی توسط محققین مدل‌های چندسطحی معرفی شد.

جدول ۱: برآورد ضرایب اثرات ثابت همراه با خطای برآورد و p-مقدار مربوطه.

متغیر	برآورد	خطای برآورد	p-مقدار
عرض از مبدأ($\gamma_{\circ\circ}$)	-۰/۰۰۵۸	۰/۰۱۷۳	۰/۷۳۹
HRS ($\gamma_{۱\circ}$)	-۰/۰۷۱۲	۰/۰۱۱۲	۰/۰۰۰
$LEAD$ ($\gamma_{۲\circ}$)	۰/۴۲۳۳	۰/۰۱۴۴	۰/۰۰۰
$G.HRS$ ($\gamma_{\circ ۱}$)	۰/۰۶۷۲	۰/۰۱۷۰	۰/۰۰۱

منابع ارزشمند در این زمینه عبارت‌اند از: ماک [۱۲]، افشارتوس [۸]، اسنایدر و بوسکر [۷] و براون و همکاران [۳].

برخلاف روش‌های به‌کار رفته قبلی در زمینه تعیین اندازه‌ی نمونه بهینه در مدل‌های چندسطحی، روش مورد استفاده در این مقاله روش زیرنمونه‌گیری است. به عبارتی دقیق‌تر ابتدا با روش زیرنمونه‌گیری نمونه‌های مختلف به دست می‌آید، آن‌گاه توان آزمون مربوطه را به‌دست می‌آوریم و سپس با کمک رابطه‌ی (۱۰) نتایج حاصل را با حالتی که تعداد کل نمونه در اختیار است، مقایسه می‌نماییم. برای دسته‌بندی مناسب نتایج نقش تأثیر تعداد نمونه‌ها بر برآورد اثرهای ثابت و تصادفی را به تفکیک بررسی می‌کنیم. بیان این نکته ضروری است که در مقاله حاضر توجه‌مان را معطوف به تأثیر تعداد گروه‌ها بر برآورد پارامترهای مدل کردیم. دلیل اصلی آن نقش کم‌رنگ تعداد زیرگروه‌ها بر عملکرد پارامترهای مدل است.

۵- الگوریتم شبیه‌سازی

فرض کنیم iامین زیرنمونه عبارت است از نمونه‌گیری بدون جایگذاری از یک مجموعه $\{X_1,\ldots,X_n\}$ با هدف تشکیل یک زیرنمونه به اندازه‌ی b، آن‌گاه تعداد زیرنمونه‌های ممکن عبارت است از:

$$C(b,n)=\binom{n}{b}$$

که معمولاً عددی بسیار بزرگ است. از این تعداد زیرنمونه تعدادی به‌صورت تصادفی انتخاب‌شده و نمونه‌ی مورد نظر برای انجام استنباط در نظر گرفته می‌شود. چنین رویکردی به انتخاب نمونه به روش زیرنمونه‌گیری معروف است [۱۴].

واضح است که روش زیرنمونه‌گیری حالت کلی‌تری از روش باز نمونه‌گیری (خودگردان‌سازی) است. یکی از تفاوت‌های زیرنمونه‌گیری با خودگردان‌سازی در روش نمونه‌گیری است؛ به این معنی که در روش زیرنمونه‌گیری، با استفاده از نمونه‌گیری بدون جایگذاری اقدام به نمونه‌گیری می‌شود ولی در روش خودگردان‌سازی با استفاده از نمونه‌گیری با جایگذاری، نمونه‌گیری انجام می‌گیرد [۱۴].

به‌طور دقیق در این مطالعه زیرنمونه‌هایی به اندازه‌ی ۵، ۱۰، ۲۰، ۴۰ و ۸۰ را در نظر گرفتیم و در ادامه با استفاده از الگوریتم زیر اقدام به شبیه‌سازی نمودیم:

الف) هر یک از زیرنمونه‌های به اندازه‌ی ۵، ۱۰، ۲۰، ۴۰ و ۸۰ را ۱۰۰ بار تکرار شود.

ب) برای هر تکرار، مدل دوسطحی به‌صورت رابطه‌ی (۱۱) به داده‌های حاصل برازش شود.

ج) میانگین برآورد پارامترها را در هر بار، به‌عنوان برآورد و انحراف استاندارد آن‌ها را خطای برآورد در نظر گرفته شود.

د) با لحاظ نمودن اطلاعات مرحله (ج) در رابطه‌ی (۹) توان آزمون را به دست آورده شود.

نتایج حاصل همراه با تفسیر آن‌ها به تفصیل در ادامه خواهد آمد. برای دسته‌بندی مناسب، نتایج را به دو دسته‌ی تأثیر اندازه‌ی نمونه بر اثرهای ثابت و تصادفی تقسیم نمودیم.

۵-۱- بررسی تأثیر اندازه‌ی نمونه بر برآورد اثرهای ثابت

اثرهای ثابت مدل عبارت هستند از: ضرایب متغیرهای *HRS*، *G.HRS* و *LEAD*. جدول ۲ شامل میانگین برآورد پارامترهای مربوط به ضرایب متغیرهای تبیینی خلق و خوی کارفرما (*LEAD*)، ساعت کاری هر فرد (*HRS*) و همچنین متوسط ساعت کاری هر گروه (*G.HRS*)، خطای برآورد و در نهایت توان آماری برای ترکیب اندازه‌ی نمونه‌های مختلف است.

جدول ۲: برآورد ضرایب اثرات ثابت، خطای برآورد و توان آزمون برای اندازه گروه‌های مختلف سطح دوم

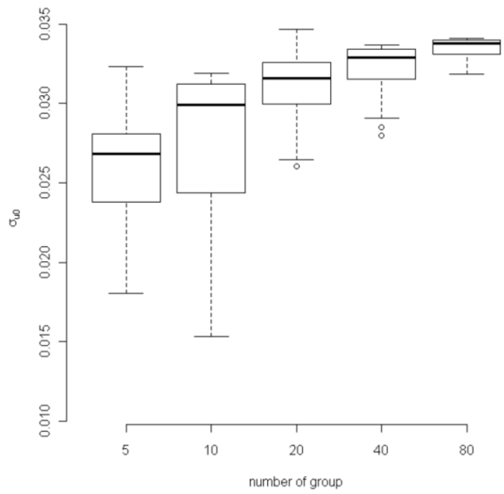
متغیر	تعداد گروه‌ها (واحدهای سطح دوم)	میانگین	خطای برآورد	توان
	۵	۰/۴۳۰۲	۰/۰۰۹۹	۰/۵۶۹۰
	۱۰	۰/۴۳۲۷	۰/۰۰۸۳	۰/۷۲۲۴
LEAD	۲۰	۰/۴۳۰۴	۰/۰۰۴۴	۰/۹۲۷۹
	۴۰	۰/۴۳۱۱	۰/۰۰۴۳	۰/۹۵۳۲
	۸۰	۰/۴۳۲۵	۰/۰۰۲۶	۰/۹۹۹۹
	۵	−۰/۰۵۷۴	۰/۰۱۱۷	۰/۸۱
	۱۰	−۰/۰۶۳۴	۰/۰۰۷۷	۰/۸۸
HRS	۲۰	−۰/۰۶۸۴	۰/۰۰۵۸	۰/۹۲
	۴۰	−۰/۰۶۸۸	۰/۰۰۴۱	۰/۹۵
	۸۰	−۰/۰۷۰۵	۰/۰۰۲۸	۰/۹۹
	۵	−۰/۰۵۰۲	۰/۰۲۵۶	۰/۸۲
	۱۰	−۰/۰۵۱۷	۰/۰۱۸۵	۰/۸۶
G.HRS	۲۰	−۰/۰۵۳۹	۰/۰۱۸۶	۰/۸۷
	۴۰	**−۰/۰۵۷۸**	**۰/۰۱۸۰**	**۰/۹۵**
	۸۰	**−۰/۰۶۳۹**	**۰/۰۱۷۷**	**۰/۹۹**

همان‌گونه که از جدول ۲ ملاحظه می‌شود با افزایش اندازه‌ی نمونه در سطح دوم توان آزمون افزایش پیدا کرده، در نهایت به یک میل می‌کند. به‌علاوه متناسب با سرعت افزایش اندازه‌ی نمونه، خطای برآورد نیز کاهش می‌یابد. این موضوع نشان‌دهنده‌ی تأثیر مستقیم تعداد واحدهای سطح دوم یا به عبارت دیگر تعداد گروه‌های مورد بررسی بر توان آزمون است. در واقع یکی از دلایل افزایش توان در مقابل افزایش تعداد واحدهای سطح دوم همین کاهش خطای برآورد است. از طرف دیگر با توجه به این‌که مقادیر حاصل از زیرنمونه‌گیری برای ارزیابی اثر پارامتر ثابت منتسب به متغیر خلق و خوی کارفرما به برآورد اثر این متغیر تحت داده‌های اصلی بسیار نزدیک و خطای برآورد آن‌ها بسیار کوچک است، می‌توان پایا بودن این اثر ثابت را تحت اندازه‌ی نمونه‌های مختلف، نتیجه گرفت. در مورد دو متغیر دیگر نیز می‌توان همین ادعا را داشت. لازم به ذکر است که متغیر چه در سطح اول و چه در سطح دوم باشد، دارای عملکردی تقریباً مشابه است. این موضوع به محقق کمک می‌کند که اگر وی به دنبال آزمونی با توان حداقل ۹۰٪ است به‌جای ۹۹ گروه می‌تواند به حداقل ۲۰ گروه اکتفا کند که این امر باعث صرفه‌جویی در وقت و هزینه خواهد شد.

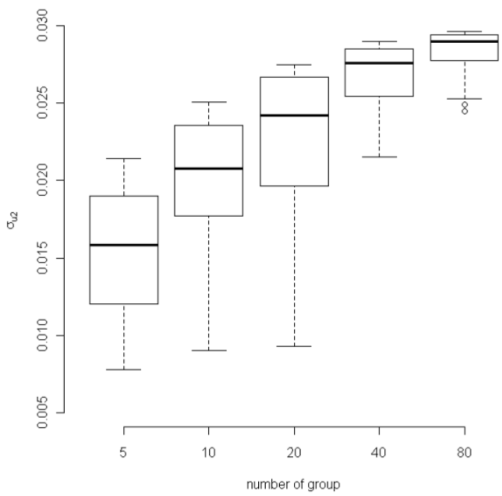
۵-۲- بررسی تأثیر اندازه‌ی نمونه بر برآورد اثرهای تصادفی

در فرآیند زیرنمونه‌گیری که در زیربخش قبل توصیف شده است، بررسی تأثیر تعداد واحدهای سطح دوم بر برآورد اثرهای تصادفی را نیز مدنظر قراردادیم. لازم به یادآوری است که اثرهای تصادفی مدل (۱۱) عبارت هستند از: $\sigma_{u_0}^2$، $\sigma_{u_2}^2$ و σ_{02} که به‌ترتیب واریانس‌های خطای نمونه‌گیری گروه‌ها منتسب به سطح دوم ناشی از تغییرات عرض از مبدأ و ضریب میانگین ساعت کاری و کوواریانس بین این دو خطا هستند.

شکل ۱، نمودار جعبه‌ای انحراف استانداردهای σ_{u_0} را نشان می‌دهد. در این شکل محور عمودی بیانگر مقدار انحراف استاندارد، محور افقی تعداد گروه‌های در نظر گرفته‌شده را برای هر زیرنمونه‌گیری نشان می‌دهد. مشاهده می‌شود که با افزایش تعداد گروه‌ها مقدار انحراف استانداردها افزایش می‌یابد و همچنین پراکندگی آن‌ها کم می‌شود. این افزایش انحراف‌ها و کاهش پراکندگی در همسایگی مقدار اصلی آن که بر اساس کل نمونه‌ها برآورد شده برابر ۰/۰۳۵ است، به حداکثر خود می‌رسد. به عبارت دیگر با انتخاب نمونه‌های کم در سطح دوم مشکل کم برآوردی اثر تصادفی σ_{u_0} پیش خواهد آمد. در نتیجه توصیه می‌شود که تعداد گروه مورد مطالعه تا حد ممکن بزرگ اختیار شود.



شکل ۱: نمودار جعبه‌ای اثرات تصادفی سطح دوم σ_{u_0}



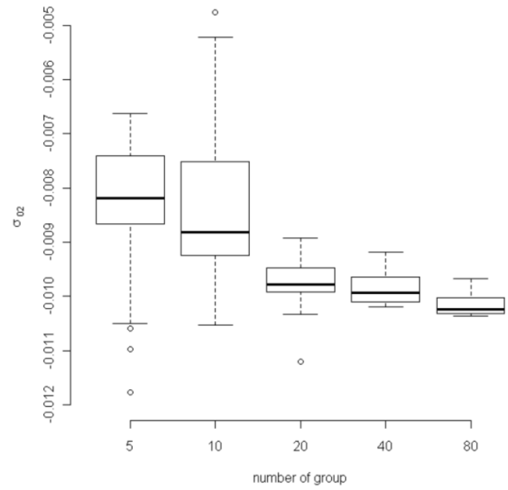
شکل ۲: نمودار جعبه‌ای اثرات تصادفی سطح دوم $\sigma_{u_۲}$ (خلق و خوی کارفرما)

شکل ۲ نمودار جعبه‌ای $\sigma_{u_۲}$ را برای زیرنمونه‌هایی از تعداد گروه‌ها نشان می‌دهد. مشابه شکل قبل محور افقی تعداد گروه‌های انتخابی و محور عمودی مقادیر انحراف استانداردهای اثر تصادفی منتسب به متغیر $LEAD$ برای ۱۰۰ بار زیرنمونه‌گیری است. همان‌گونه که ملاحظه می‌شود با افزایش تعداد گروه‌ها پراکندگی انحراف استانداردها کاهش و نهایتا به مقدار واقعی برآورد $\sigma_{u_۲}$ که برابر ۰/۰۳۱ است، نزدیک‌تر می‌شوند. واضح است که با اندازه‌ی نمونه‌ی کم از سطوح دوم مشکل کم برآوردی برآورد اثر تصادفی منتسب به متغیر تبیینی سطح دوم رخ می-دهد و این وضعیت حتی حادتر از اثر تصادفی سطح اول است.

شکل ۳ نمودار جعبه‌ای را برای برآورد کوواریانس بین $u_{۰j}$ و $u_{۲j}$ بر اساس روش زیرنمونه‌گیری از تعداد گروه‌ها نشان می‌دهد. ملاحظه می‌شود که با اندازه‌ی نمونه‌ی کم در سطح دوم ابتدا مشکل بیش‌برآوردی اتفاق می‌افتد. اما با افزایش نمونه، برآورد کوواریانس به مقدار واقعی نزدیک‌تر می‌شود. از این‌رو توصیه می‌شود برای رسیدن به برآورد مناسبی از کوواریانس بین خطاهای منتسب به متغیر توضیحی اندازه‌ی نمونه اختیاری منتسب به سطح دوم تا حدودی بزرگ اختیار شود.

شکل ۳ نمودار جعبه‌ای را برای برآورد کوواریانس بین $u_{۰j}$ و $u_{۲j}$ بر اساس روش زیرنمونه‌گیری از تعداد گروه‌ها نشان می‌دهد. ملاحظه می‌شود که با اندازه‌ی نمونه‌ی کم در سطح دوم ابتدا مشکل بیش‌برآوردی اتفاق می‌افتد. اما با افزایش نمونه، برآورد کوواریانس به مقدار واقعی نزدیک‌تر می‌شود. از این‌رو توصیه می‌شود برای رسیدن به برآورد مناسبی از

کوواریانس بین خطاهای منتسب به متغیر توضیحی اندازه‌ی نمونه اختیاری منتسب به سطح دوم تا حدودی بزرگ اختیار شود.



شکل ۳: نمودار جعبه‌ای اثرات تصادفی سطح دوم σ_{02} (کوواریانس بین u_{0j} و u_{2j})

۶- بحث و نتیجه‌گیری

مدل‌های چندسطحی یکی از کاربردی‌ترین مدل‌ها برای بررسی رفتار داده‌ها با ساختار همبستگی درون‌گروهی هستند. تعیین اندازه‌ی نمونه در مدل‌های چندسطحی پیش از برازش چنین مدل‌هایی به محقق کمک می‌کند تا برآوردهای دقیق‌تر و نتایج معتبرتری را به‌دست آوریم. در بررسی تأثیر اندازه نمونه بر برآورد پارامترهای مدل‌های دوسطحی، افشارتوس در [۸] به ارزیابی وجود اریبی و مقدار واریانس برآوردگر پرداخته در حالی که ماک [۱۲] بیش‌برآوردی و کم‌برآوردی برآوردگرها را در مقابل تغییر متقابل اندازه نمونه در سطح اول و دوم مد نظر قرارداده است. بررسی جنبه‌های نظری اندازه نمونه‌های بهینه برای پاسخ‌های نرمال با تاکید بر کارائی واریانس برآوردگر توسط اسنایدر در [۷] انجام شد. فرموله‌بندی بخشی از این فعالیت‌ها با الهام از الگوریتم‌های شبیه‌سازی در [۳] آمده است. موضوع زیرنمونه‌گیری رویکرد جدیدی به این موضوع است که نویسندگان مقاله حاضر به بررسی آن پرداخته‌اند. با این حال، نتایج حاصل از این تحقیق تا حدود زیادی مشابه یافته‌های محققین قبلی بوده است. یکی از روش‌های ارزیابی تعیین اندازه‌ی نمونه بررسی تأثیر تعداد واحدهای سطح دوم روی توان و برآورد اثرهای ثابت و تصادفی با استفاده از روش زیرنمونه‌گیری است. نتایج شبیه‌سازی حاصل از این مقاله نشان داد که برای دستیابی به توان و دقت بیشتر بهتر است تعداد واحدهای سطح دوم تا حد

ممکن بیشتر از سطح اول اختیار شود. واضح است که محدودیت‌های دیگر از قبیل وقت، هزینه و تعداد گروه‌های در دسترس (تعداد واحدهای سطح دوم) نیز تأثیر بسزایی بر تعیین اندازه‌ی نمونه خواهد داشت. لحاظ نمودن این عوامل همراه با اعمال رویکردهای بیزی موضوع تحقیقات آتی نویسندگان این مقاله خواهد بود.

تقدیر و تشکر

نویسندگان مراتب قدردانی خود را از داوران محترم برای ارائه نقطه نظراتشان و هیئت تحریریه-ی محترم مجله اعلام می‌دارند.

مراجع

[1] Cohen, M. (1998), Determining Sample Size for Surveys with Data Analyzed by Hierarchical Linear Models, *Journal of Official Statistics*, **14**, 267-257.

[2] Goldstein, H. (2010), *Multilevel Statistical Models*, (4th ed), Chichester: John Wiley and Sons.

[3] Browne, W. J, Golalizadeh, M. and Parker, R. (2009), *A Guide to Sample Size Calculations for Random Effect Models via Simulation and MLPowSim Software Package*, Bristol University Press, Bristol.

[4] Longford, N.T. (1987), A First Scoring Algorithm for Maximum Likelihood Estimation in Unbalanced Mixed Models with Nested Effects, *Biometrika*, **74**, 812-827.

[5] Goldstein, H. and Silver, R. (1989), *Multilevel and Multivariate Models in Survey Analysis*, In Skinner, C. J, Holt, D. and Smith, T. M. F (ed.), Analysis of Complex Surveys, 221-235. New York: John Wiley and Sons.

[6] Snijder, T.A.B. and Bosker, R.J. (1993), Standard Errors and Sample Size for Two-Level Research, *Journal of Educational Statistics*, **18**, 237-259.

[7] Snijder, T.A.B. and Bosker, R.J. (1999), *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling Multilevel StatisticalModels*, Sage Publications: London.

[8] Afshartous, D. (1995), Determination of Sample Size for Multilevel Model Design, *Paper Presented at the AERA meeting in San Francisco*, CA.

[9] Gelman, A. and Hill, J. (2007), *Data Analysis Using Regression and Multilevel/ Hierarchical Models*, Cambridge: Cambridge University Press.

[10] Twisk, J.W.R. (2006). *Applied Multilevel Analysis: Practical Guides to Biostatisticals and Epidemiology*. Cambridge: University Press.

[11] Cochran, W.G. (1977), *Sampling Techniques*, (3rd ed). New York: John Wiley and Sons.

[12] Mok, M. (1995), Sample Size Requirements for 2-Level Design in Education Research, *Multilevel Modeling Newsletter*, **7**, 11-15.

[13] Bliese, P.D. and Halverson, R.R. (1998), Group Size and Measures of Group-Level Properties: An Examination of Eta-Squared and ICC Value, *Journal of Management,* **24**, 157-172.

[14] Politis, D.N., Romano, J.P. and Wolf, M. (1999), *Subsampling*, New York: Springer.

The Effects of Sample Size in Multilevel Models via Sub-sampling Approach

Omid Akhgari and Mousa Golalizadeh

Department of Statistics, Tarbiat Modares University, Tehran, Iran

Abstract

There are numerous research topics in different fields of study including social, medical, and agricultural sciences which contain data having intra-class correlation structure. Obviously, due to ignoring the underlying correlation structure, simple linear regression models do not have acceptable efficiency for this data. For this situation, the suitable models, which have the ability to include the within groups correlation and between groups independency, are known as multilevel models. Determining appropriate sample size at different levels of hierarchy is the main objective in studying these models. Due to high complexities, rather than investigating any particular sampling method to select the sample size, the effect of different sample size on performance of the estimators are usually considered. In the present study, following sub-sampling method, the influence of different sizes of the sample at the first and second levels of hierarchy on estimating the fixed and random effects are investigated. Furthermore, due to the close relationship between determining sample size and the power of the statistical test, and also because of the impact of other components such as design effect and significance level, the performance of different sampling combinations are evaluated. Results of our investigation show that the increasing sample size at the second level of a two-level model leads to high power on tests about the fixed and random effects.

Keywords: Multilevel model, Optimal Sample size, Fixed effects, Random effects, Sub-sampling, Statistical power.

Mathematics Subject Classification (2000): 62J12, 62K05.