تعیین حجم نمونه با کمترین هزینه برای دو جامعه به روش بیز با استفاده از توابع زیان متقارن و نامتقارن

مسعود قاسمی بهجانی - میلاد اسدزاده
گروه آمار، دانشگاه سیستان و بلوچستان - گروه آمار، دانشگاه یاسوج

چکیده

در این مقاله با استفاده از توابع زیان درجه دوم و لینکس، حجم نمونه را به روش بیز در برآورد تفاضل بین میانگین‌های دو توزیع نرمال از دو جامعه‌ی متفاوت به قسمی می‌یابیم که هزینه‌ی نمونه‌گیری مینیمم شود. در این فرآیند برای مینیمم کردن هزینه از تابع هزینه‌ی لیندلی استفاده می‌شود. در برخی حالات بدلیل دشوار بودن محاسبات، حجم نمونه را نمی‌توان به روش تحلیلی بدست آورد، به‌همین دلیل به کمک روش‌های عددی حجم نمونه‌ی مطلوب را بدست می‌آوریم.

واژه‌های کلیدی: توزیع نرمال، تفاضل میانگین‌ها، تابع هزینه و ریسک پسین.

۱ مقدمه

محاسبه‌ی حجم نمونه در مسایل عملی سوال مهمی است که مستقیماً با هزینه‌ی مطالعه و گردآوری داده‌ها ارتباط دارد. در هر پژوهش، دانستن تعداد مشاهدات (اندازه‌ی نمونه) به‌طوری‌که هزینه‌ی کلی طرح نیز مینیمم شود، مساله‌ای است که از اهمیت زیادی برخوردار است. هزینه‌ی کلی، شامل زیان حاصل شده از تصمیم‌گیری و هزینه‌ی آنالیز و هدایت آزمایش‌های انجام شده است. این هزینه به عوامل زیادی بستگی دارد. از جمله‌ی عوامل مهم، یکی اندازه‌ی نمونه‌ی نهایی و دیگری روش جمع‌آوری داده‌ها است. برای داشتن دقت بالاتر و در نتیجه افزایش سود باید آزمایش‌های بیشتری انجام شود، که این امر باعث افزایش هزینه‌ی آزمایش‌ها می‌شود. اولین مقاله بر اساس نظریه‌ی تصمیم برای تعیین اندازه‌ی نمونه توسط گروندی[۱] و همکاران [۶] بر پایه‌ی تابع زیان ارائه شد، این روش توسط همیلتون[۲] [۷] دوباره مطرح شد و مورد توجه بسیاری از پژوهشگران قرار گرفت. مسئله‌ی پیدا کردن اندازه‌ی نمونه‌ی بهینه برای هر دو نوع نمونه‌گیری ثابت و دنباله‌ای توسط برگر[۳] [۳] بحث شده است. برای برآورد میانگین یک توزیع

[۱] Grundy
[۲] Hamilton
[۳] Berger

اختیاری، گلدستاین[۴] [۵] یک معیار بیزی را بر اساس تغییر مورد انتظار برآورد نقطه‌ای برای میانگین مقادیر نمونه‌ی بعدی مطرح کرده، و حد بالایی را برای اندازه‌ی نمونه محاسبه نموده است. آدکوک[۵] [۱] مطالعه‌ی نسبتا جامعی روی هر دو دسته از روش‌ها ارائه کرده است. فام گیا و ترکان[۶] [۱۳] مسئله را براساس مقادیر مورد انتظار از اطلاعات نمونه بررسی کرده‌اند. استالارد [۷] [۱۷] با به‌کارگیری این روش در فاز دوم آزمایه‌های بالینی، اندازه‌ی نمونه را با ماکسیمم‌سازی تابع سود خاصی بدست آورد. پزشک و گیتینز [۸] [۱۲] مساله‌ی تعیین اندازه‌ی نمونه را با استفاده از روش بیز برای یک آزمایه‌ی بالینی مطرح و آن را برای حالتی خاص حل کردند. اوهاگان و استیونس[۹] [۱۰] یک تابع هدف شامل هزینه و سود انجام یک آزمایش تصادفی را مطرح نموده و با بهینه‌سازی آن، اندازه‌ی نمونه‌ی آزمایش تصادفی را به روش بیزی به دست آورند. همچنین پزشک [۱۱]، مروری بر روش‌های مختلف بیزی برای تعیین اندازه‌ی نمونه‌ی آزمایش‌های تصادفی ارائه داد. ساهو و اسمیت [۱۰] [۱۵] در مورد تعیین حجم نمونه در آزمایه‌های بالینی و حسابرسی مالی بحث کردند. در طول دهه‌ی اخیر، تعیین اندازه‌ی نمونه از نقطه نظر بیزی مورد توجه بسیاری از پژوهشگران، صنعت و دولت قرار گرفته است. اگرچه هنوز بین دانشمندان بیزی و کلاسیک بحث و اختلاف نظر وجود دارد، برگر، بوکای[۱۱] و وانگ[۱۲] [۲، ۴] به‌صورت موفقیت‌آمیزی، شایستگی‌های هردو نگرش کلاسیک و بیزی را بیان کردند. به هر حال، هیچ یک از بحث‌ها راجع به تعیین حجم نمونه از نقطه‌نظر بیزی برای آزمایش‌های بالینی در مرحله‌ی طرح‌ریزی میسر نشد، بنابراین بسیاری از محققان از شبیه سازی استفاده نمودند. در این مقاله با در نظر گرفتن تابع هزینه‌ی لیندلی و تابع سود به روش بیز به دنبال تعیین حجم نمونه می‌باشیم به طوری که هزینه‌ی نمونه‌گیری مینیمم شود.

[۴] Goldstein
[۵] Adcock
[۶] Pham-Gia and Turkkan
[۷] Stalard
[۸] Pezeshk and Gitins
[۹] O'hagan and Stivence
[۱۰] Sahu and Smith
[۱۱] Boukai
[۱۲] Wang

۱.۱ توابع زیان و ریسک پسین

اگر δ برآورد θ باشد، تابع زیان متقارن درجه دوم بصورت زیر تعریف می‌شود:

$$L(\delta,\theta) = (\delta-\theta)^{2}.$$

فرض کنید $p(\theta|\underline{x})$ تابع چگالی پسین باشد در این صورت ریسک پسین بر اساس تابع زیان درجه دوم بصورت زیر است:

$$\begin{aligned} PR &= \int L(\delta,\theta)p(\theta|\underline{x})d\theta, \\ &= Var(\theta|\underline{X}), \end{aligned} \qquad (1)$$

که همان واریانس پسین است. زلنر[۱۳] [۱۸] تابع زیان لینکس را بصورت زیر معرفی کرد:

$$L(\delta,\theta) = \exp(b(\delta-\theta)) - b(\delta-\theta) - 1.$$

[۱۳] Zellner

برآورد بیز بر اساس این تابع زیان بصورت زیر است:

$$\delta = -\frac{1}{b} \ln E\left[\exp(-b\theta|\underline{X})\right]. \qquad (۲)$$

ریسک پسین بر اساس تابع زیان لینکس عبارت است از:

$$PR = b(m - \delta), \qquad (۳)$$

که b پارامتر شکل و m میانگین پسین است.

۲.۱ تابع هزینه

لیندلی[۱۴] [۸] با فرض اینکه c_0 هزینه‌ی زیرساختی از نمونه‌گیری یا هر هزینه‌ی دیگر در نمونه‌گیری باشد و c هزینه‌ی هر واحد نمونه‌گیری در نظر گرفته شود تابع هزینه برای حجم نمونه را به‌صورت

$$C(n) = c_0 + cn; \quad n > 0, \qquad (۴)$$

معرفی کرد. سیف الاسلام[۱۵] [۱۶] با اضافه کردن ریسک پسین برآوردگر بیز به این تابع هزینه، تابع هزینه‌ی کل را به‌صورت

$$TC(n) = c_0 + cn + PR,$$

بدست آورد. اگر بخواهیم نمونه‌ها را از دو جامعه انتخاب کنیم باید تابع هزینه را به شکلی که در ادامه گفته می‌شود در نظر بگیریم. فرض کنید c_1 هزینه‌ی نمونه‌گیری برای هر واحد از نمونه‌ای با حجم n_1 از یک جامعه و c_2 هزینه‌ی نمونه‌گیری برای هر واحد از نمونه‌ای با حجم n_2 از جامعه‌ی دیگر باشد. تابع هزینه بصورت زیر بدست می‌آید:

$$C(n) = c_0 + c_1 n_1 + c_2 n_2. \qquad (۵)$$

بنابراین هزینه‌ی کل بصورت زیر می‌باشد:

$$TC(n) = c_0 + c_1 n_1 + c_2 n_2 + PR.$$

۲ تعیین حجم نمونه

در این بخش حجم نمونه‌ی مطلوب در برآورد تفاضل بین میانگین‌های دو توزیع نرمال از دو جامعه‌ی متفاوت را بر اساس توابع زیان درجه دوم و لینکس بدست می‌آوریم.

فرض کنید $X_1, X_2, ..., X_{n_1}$ یک نمونه‌ی تصادفی به حجم n_1 با میانگین μ_1 و دقت $\xi_1(=\frac{1}{\sigma_1^2})$ و $Y_1, Y_2, ..., Y_{n_2}$ یک نمونه‌ی تصادفی دیگر به حجم n_2 با میانگین μ_2 و دقت $\xi_2(=\frac{1}{\sigma_2^2})$ از توزیع نرمال باشد. توزیع‌های مزدوج پیشین را برای μ_1، $N(\mu_{01}, n_{01}\xi_1)$ و برای μ_2، $N(\mu_{02}, n_{02}\xi_2)$ در نظر می‌گیریم که n_0 حجم نمونه‌ی پیشین است. می‌خواهیم حجم نمونه‌ی مطلوب در برآورد تفاضل میانگین‌های توزیع نرمال، θ که $\theta = \mu_1 - \mu_2$ است را بر اساس زیان درجه دوم و زیان لینکس بدست آوریم. در ابتدا فرض می‌کنیم دو جامعه دارای دقت‌های برابر باشند یعنی $\xi = \xi_1 = \xi_2$. سپس با شرط $n_1 \neq n_2$، فرض می‌کنیم دقت‌ها معلوم اما نابرابرند، بعبارت دیگر $\xi_1 \neq \xi_2$.

[۱۴] Lindley
[۱۵] Saiful Islam

۱.۲ تعیین حجم نمونه وقتی دقت‌های دو جامعه معلوم و برابر باشند

چگالی پسین θ برای داده‌های $\underline{X}$ و $\underline{Y}$ بصورت زیر می‌باشد:

$$\theta|\underline{x},\underline{y} \sim N\left\{m_2 - m_1, \frac{\xi(n_{01}+n_1)(n_{02}+n_2)}{n_1+n_2+n_{01}+n_{02}}\right\}, \qquad (۶)$$

که m_1 و m_2 بترتیب برآوردهای μ_1 و μ_2 هستند. فرض کنید $m = m_2 - m_1$ میانگین پسین باشد به طوری $m_1 = \frac{n_{01}\mu_{01}+n_1\bar{x}}{n_{01}+n_1}$ و $m_2 = \frac{n_{02}\mu_{02}+n_2\bar{y}}{n_{02}+n_2}$.

۱.۱.۲ تعیین حجم نمونه بر اساس زیان درجه دوم

در ابتدا حجم نمونه‌ی مطلوب را بر اساس زیان درجه دوم بدست می‌آوریم. طبق رابطه‌ی (۱)، ریسک پسین بر اساس تابع زیان درجه دوم، واریانس پسین است. از طرفی بر اساس رابطه‌ی (۶)، واریانس پسین برابر است با: $\frac{n_1+n_2+n_{01}+n_{02}}{\xi(n_{01}+n_1)(n_{02}+n_2)}$. حال با اضافه کردن ریسک پسین به تابع هزینه‌ی (۵)، هزینه‌ی کل به صورت زیر بدست می‌آید:

$$TC(n) = c_0 + c_1 n_1 + c_2 n_2 + \frac{n_1+n_2+n_{01}+n_{02}}{\xi(n_{01}+n_1)(n_{02}+n_2)}. \qquad (۷)$$

در ادامه با در نظر گرفتن حالتی که حجم هر دو نمونه برابر باشد، حجم نمونه‌های پیشین برابر باشند اما هزینه‌ی نمونه‌گیری برای هر واحد برای دو نمونه متفاوت باشد، به عبارت دیگر $n_1 = n_2 = n$، $n_{01} = n_{02} = n_0$ و $c_1 \neq c_2$، حجم نمونه‌ی مطلوب را بدست می‌آوریم. با این تفاسیر رابطه‌ی (۷) را می‌توان بصورت زیر نوشت:

$$TC(n) = c_0 + (c_1 + c_2)n + \frac{2}{\xi(n_0 + n)}. \qquad (۸)$$

حال برای بدست آوردن حجم نمونه‌ی مطلوب از رابطه‌ی (۸) نسبت به n مشتق گرفته و برابر با صفر قرار می‌دهیم و در نتیجه داریم:

$$n^* = \sqrt{\frac{2}{\xi(c_1 + c_2)}} - n_0,$$

و در صورتی که هزینه‌ی نمونه‌گیری برای هر واحد در دو جامعه برابر باشد حجم نمونه‌ی مطلوب بصورت زیر است:

$$n^* = \frac{1}{\sqrt{c\xi}} - n_0.$$

حال با فرض $n_1 = n_2 = n$، $n_{01} \neq n_{02}$ و $c_1 \neq c_2$ حجم نمونه‌ی مطلوب را برای هر یک از جوامع بدست می‌آوریم. در ابتدا طبق رابطه‌ی (۷) داریم:

$$TC(n) = c_0 + (c_1 + c_2)n + \frac{2n + n_{01} + n_{02}}{\xi(n_{01} + n)(n_{02} + n)}. \qquad (۹)$$

حال از رابطه‌ی (۹) نسبت به n مشتق گرفته و برابر با صفر قرار می‌دهیم:

$$An^4 + Bn^3 + Cn^2 + Dn + E = 0, \qquad (۱۰)$$

که

$$A = c_1 + c_2, B = 2AN, C = AN^2 + 2AN' - 2k,$$
$$D = 2ANN' - 2kN, E = AN'^2 + 2kN' - N^2$$
$$N = n_{01} + n_{02}, N' = n_{01}n_{02}, k = \frac{1}{\xi}.$$

حال می‌توان به سادگی رابطه‌ی (۱۰) را برای تعیین حجم نمونه به ازای مقادیر معلوم n_{01}، n_{02}، c_1 و c_2 حل کرد. در اینجا هزینه‌ی نمونه‌گیری برای هر واحد در دو جامعه برابر است اما بین جوامع تفاوت وجود دارد. برای حل عددی این معادله فرض می‌کنیم $n_{01} = n_{02} = 10$ و $\xi = 1$. این معادله‌ی چند جمله‌ای را به دلیلی پیچیدگی با استفاده از $Maple$ [۹] و فرض $c_1 + c_2 = c$ حل می‌کنیم. در فرآیند حل معادله به چهار ریشه از n برخورد می‌کنیم که سه ریشه منفی و فقط یکی مثبت است و همان ریشه‌ی مثبت را بعنوان حجم نمونه در نظر می‌گیریم زیرا حجم نمونه هیچگاه منفی نمی‌شود. در شکل (۱) میبینیم که با افزایش هزینه‌ی نمونه‌گیری، حجم نمونه‌ی مطلوب کاهش می‌یابد.

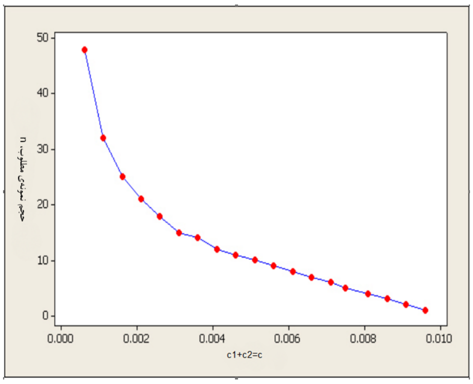
حال حجم نمونه‌ی مطلوب را با شرط $n_1 \neq n_2$، $n_{01} \neq n_{02}$، $c_1 \neq c_2$ و $\xi_1 = \xi_2 = \xi$ بدست می‌آوریم. بدین منظور ابتدا از رابطه‌ی (۷) نسبت به n_1 مشتق گرفته و برابر با صفر قرار می‌دهیم و در نتیجه داریم:

$$n_1^* = \frac{1}{\sqrt{c_1\xi}} - n_{01}.$$

سپس همین فرآیند را برای n_2 تکرار می‌کنیم:

$$n_1^* = \frac{1}{\sqrt{c_2\xi}} - n_{02}.$$

حال یک زوج از حجم نمونه‌ی مطلوب (n_1^*, n_2^*) داریم که هزینه‌ی کل در رابطه‌ی (۷) را مینیمم می‌کند.



شکل ۱: حجم نمونه‌ی مطلوب به ازای مقادیر مختلف c.

۲.۱.۲ تعیین حجم نمونه بر اساس زیان لینکس

در بخش حجم نمونه‌ی مطلوب را بر اساس تابع زیان لینکس بدست می‌آوریم. همانطور که گفته شد چگالی پسین برای θ بصورت رابطه‌ی (۶) است . بنابراین طبق رابطه‌ی (۲)، برآورد بیز بر اساس تابع زیان لینکس عبارت است از:

$$\delta = m - \frac{b\nu^2}{2}, \tag{11}$$

که ν^2 واریانس پسین است. در نتیجه طبق رابطه‌ی (۳)، ریسک پسین به صورت زیر است:

$$PR = \frac{b^2\nu^2}{2}.$$

با اضافه کردن ریسک پسین به تابع هزینه‌ی (۵)، هزینه‌ی کل بصورت زیر بدست می‌آید:

$$TC(n) = c_0 + c_1n_1 + c_2n_2 + \frac{b^2(n_1 + n_2 + n_{01} + n_{02})}{2\xi(n_{01} + n_1)(n_{02} + n_2)}. \tag{12}$$

حال با فرض $n_{01} = n_{02} = n_0$ و $n_1 = n_2 = n$ حجم نمونه‌ی مطلوب را بدست می‌آوریم. بدین منظور از رابطه‌ی (۱۲) نسبت به n مشتق گرفته و برابر با صفر قرار می‌دهیم. بنابراین داریم:

$$n^* = b\sqrt{\frac{1}{\xi(c_1 + c_2)}} - n_0,$$

و اگر هزینه‌ی نمونه‌گیری برای هر واحد در هر دو جامعه برابر باشد داریم:

$$n^* = b\sqrt{\frac{1}{2\xi c}} - n_0.$$

از طرفی با فرض $n_1 = n_2 = n$، $n_{01} \neq n_{02}$ و $c_1 \neq c_2$ تابع هزینه‌ی کل بهصورت زیر بدست می‌آید:

$$TC(n) = c_0 + (c_1 + c_2)n + \frac{b^2(2n + n_{01} + n_{02})}{2\xi(n_{01} + n)(n_{02} + n)}. \qquad (13)$$

حال برای بدست آوردن حجم نمونه‌ی مطلوب از رابطه‌ی (۱۳) نسبت به n مشتق گرفته و برابر با صفر قرار می‌دهیم:

$$An^4 + Bn^3 + Cn^2 + Dn + E = 0, \qquad (14)$$

که

$$A = c_1 + c_2, B = 2AN, C = AN^2 + 2AN' - 2k,$$
$$D = 2ANN_0' - 2kN_0, E = AN'^2 + 2kN' - N^2$$
$$N = n_{01} + n_{02}, N' = n_{01}n_{02}, k = \frac{b^2}{2\xi}.$$

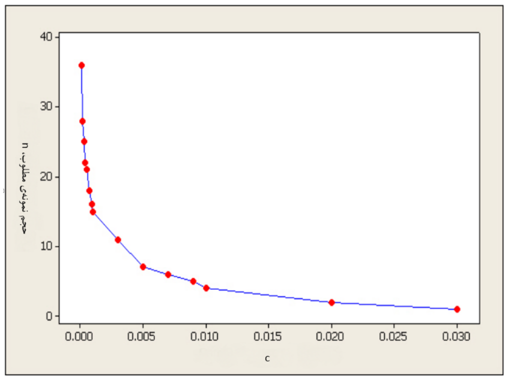
برای مقادیر معلوم c_1، c_2، n_{01}، n_{02}، ξ و b به راحتی می‌توان معادله‌ی (۱۴) را حل کرد و حجم نمونه‌ی مطلوب را بدست آورد. حال با فرض $n_1 \neq n_2$، $n_{01} \neq n_{02}$ و $c_1 \neq c_2$، می‌توان حجم نمونه‌ی مطلوب را بطور جداگانه برای n_1 و n_2 بدست آورد. بدین منظور از رابطه‌ی (۱۲) بطور جداگانه نسبت به n_1 و n_2 مشتق گرفته و برابر با صفر قرار می‌دهیم و در نتیجه داریم:

$$n_1^* = \frac{b}{\sqrt{2\xi c_1}} - n_{01}, n_2^* = \frac{b}{\sqrt{2\xi c_2}} - n_{02}.$$

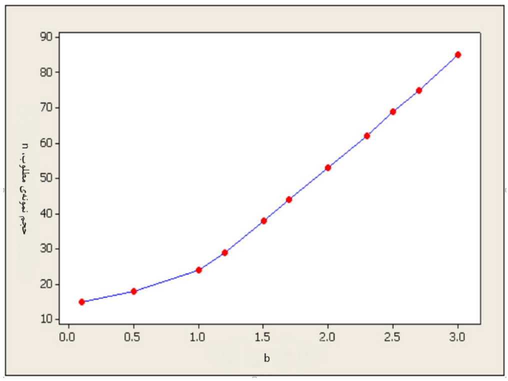
برای مقادیر معلوم c_1، c_2، n_{01}، n_{02}، ξ و b می‌توان حجم نمونه‌ی مطلوب را بدست آورد که n_1^* و n_2^* با هم هزینه‌ی کل در رابطه‌ی (۱۲) را مینیمم می‌کنند. معادله‌ی (۱۴) را بدلیل پیچیدگی با استفاده از نرم افزار $Maple$ حل می‌کنیم و در نتیجه به چهار ریشه برخورد می‌کنیم که فقط یک ریشه مثبت است و همان ریشه‌ی مثبت حجم نمونه‌ی مطلوب می‌باشد. نمودار حجم نمونه‌ی مطلوب به ازای مقادیر مختلف هزینه‌ی نمونه‌گیری برای هر واحد در شکل (۲) رسم شده است و مشاهده می‌شود که اگر هزینه‌ی نمونه‌گیری برای هر واحد افزایش یابد، حجم نمونه‌ی مطلوب کاهش می‌یابد. در شکل (۳) مشاهده می‌شود که با افزایش b حجم نمونه‌ی مطلوب کاهش می‌یابد.

۲.۲ تعیین حجم نمونه وقتی دقت‌های دو جامعه معلوم اما نابرابرند

حال حجم نمونه‌ی مطلوب را در برآورد تفاضل میانگین‌های دو توزیع نرمال بوسیله‌ی $\theta = \mu_1 - \mu_2$ بر اساس تابع زیان درجه دوم و با شرط $\xi_1 \neq \xi_2$ بدست می‌آوریم. در اینجا چگالی پسین θ بصورت زیر می‌باشد:



شکل ۲: حجم نمونه‌ی مطلوب به ازای مقادیر مختلف c، با شرط $\xi = 1$، $n_{01} = 8$، $n_{02} = 12$ و $b = 0/1$.



شکل ۳: حجم نمونه‌ی مطلوب به ازای مقادیر مختلف b، با شرط $\xi = 1$، $n_{01} = 8$، $n_{02} = 12$ و $c = 0/001$.

(۱۵)
$$\theta|\underline{x},\underline{y} \sim N\left\{m_2 - m_1, \frac{\xi_1\xi_2(n_{01}+n_1)(n_{02}+n_2)}{\xi_1(n_1+n_{01})+\xi_2(n_2++n_{02})}\right\},$$

که $m_1 = \frac{n_{01}\mu_{01}+n_1\bar{x}}{n_{01}+n_1}$ و $m_2 = \frac{n_{02}\mu_{02}+n_2\bar{y}}{n_{02}+n_2}$.

۱.۲.۲ تعیین حجم نمونه بر اساس تابع زیان تابع درجه دوم

در ابتدا حجم نمونه‌ی مطلوب را بر اساس تابع زیان درجه دوم بدست می‌آوریم. طبق رابطه‌ی (۱) ریسک پسین بر اساس تابع زیان درجه دوم واریانس پسین است. از طرفی بر اساس رابطه‌ی (۱۵) واریانس پسین بصورت $\frac{\xi_1(n_1+n_{01})+\xi_2(n_2+n_{02})}{\xi_1\xi_2(n_{01}+n_1)(n_{02}+n_2)}$ می‌باشد. حال با اضافه کردن ریسک پسین به تابع هزینه‌ی (۵)، هزینه‌ی کل عبارت است از:

(۱۶)
$$TC(n) = c_0 + c_1n_1 + c_2n_2 + \frac{\xi_1(n_1+n_{01})+\xi_2(n_2+n_{02})}{\xi_1\xi_2(n_{01}+n_1)(n_{02}+n_2)}.$$

حال برای پیدا کردن حجم نمونه‌ی مطلوب (n_1^*, n_2^*)، از رابطه‌ی (۱۶) بطور جداگانه نسبت به n_1 و n_2 مشتق گرفته و برابر با صفر قرار می‌دهیم. در نتیجه داریم:

$$n_1^* = \frac{1}{\sqrt{c_1\xi_1}} - n_{01}, n_2^* = \frac{1}{\sqrt{c_2\xi_2}} - n_{02}.$$

که برای مقادیر معلوم n_{01}، n_{02}، c_1، c_2 ، ξ_1 و ξ_2، دوتایی (n_1^*, n_2^*) هزینه‌ی کل در رابطه‌ی (۱۶) را مینیمم می‌کند.

۲.۲.۲ تعیین حجم نمونه بر اساس تابع زیان لینکس

حال حجم نمونه‌ی مطلوب را بر اساس تابع زیان لینکس بدست می‌آوریم. براساس رابطه‌ی (۳)، در اینجا ریسک پسین بصورت زیر می‌باشد:

$$PR = \frac{b^2}{2} \times \frac{\xi_1(n_1+n_{01})+\xi_2(n_2+n_{02})}{\xi_1\xi_2(n_1+n_{01})(n_2+n_{02})}.$$

با اضافه کردن ریسک پسین به تابع هزینه‌ی (۵)، هزینه‌ی کل بصورت زیر بدست می‌آید:

(۱۷)
$$TC(n) = c_0+c_1n_1+c_2n_2+\frac{b^2}{2}\times\frac{\xi_1(n_1+n_{01})+\xi_2(n_2+n_{02})}{\xi_1\xi_2(n_1+n_{01})(n_2+n_{02})}.$$

بنابراین برای بدست آوردن حجم نمونه‌ی مطلوب از رابطه‌ی (۱۷) بطور جداگانه نسبت به n_1 و n_2 مشتق

گرفته و برابر با صفر قرار می‌دهیم و در نتیجه داریم:

$$n_1^* = b\sqrt{\frac{1}{2c_1\xi_1}} - n_{01}, n_2^* = b\sqrt{\frac{1}{2c_2\xi_2}} - n_{02}.$$

برای مقادیر معلوم c_1، c_2، n_{01}، n_{02}، ξ و b به راحتی می‌توان حجم نمونه‌ی مطلوب را بدست آورد. زوج (n_1^*, n_2^*)، هزینه‌ی کل در رابطه‌ی (۱۷) را مینیمم می‌کند.

۳ نتیجه‌گیری

در این مقاله حجم نمونه‌ی مطلوب را در برآورد تفاضل میانگین‌های دو توزیع نرمال با شرط معلوم بودن دقت بر اساس توابع زیان متقارن درجه دوم و نامتقارن لینکس محاسبه کردیم. در بعضی موارد بدلیل پیچیدگی از نرم افزار برای حل معادلات استفاده شد. همانطورکه مشاهده می‌شود، در همه‌ی حالات هزینه‌ی نمونه‌گیری برای هر واحد با حجم نمونه‌ی مطلوب رابطه‌ی عکس اما مقدار پارامتر شکل، b با حجم نمونه‌ی مطلوب رابطه‌ی مستقیم دارد. بعبارت دیگر، با افزایش هزینه‌ی نمونه‌گیری برای هر واحد، حجم نمونه‌ی مطلوب کاهش یافته، در صورتی که با افزایش b حجم نمونه‌ی مطلوب نیز افزایش می‌یابد.

مراجع

[1] Adcock, C. J. (1988). A Bayesian approach to calculating Sample size. The Statistician, 37, 433-439.

[2] Berger, J. (1985). Statistical Decision Theory and Bayesian Analysis, 2nd edn. New York: Springer.

[3] Berger, J.O. (1980). Statistical Decision Theory - Foundation Concepts and Methods, Springer - Verlag.

[4] Berger, J. O., Boukai, B. and Wang, Y. (1997). Unified frequentist and Bayesian testing of a precise hypothesis. Statist. Sci, 12, 133-160.

[5] Goldstein, M. (1981). A Bayesian Criterion for Sample Size. Ann. Statist, 9, 670-672.

[6] Grundy, P. M., Healy, M. J. R. and Rees, D. H. (1956). Economic choice of the amount of experimentation. J. R. Statist. Soc. A, 18, 32-48.

[7] Hamilton, H. W. (1968). Confience Properties of Bayesian Interval Estimates. Journal of the Royal Statistical Society, Series B: Methodological, 30, 535-544.

[8] Lindley, D. V. (1972). Bayesian Statistics, a Review. Philadelphia: Society for Industrial and Applied Mathematics.

[9] Maple 13 (1996-2009) Copyright ©Maplesoft, a division of Waterloo Maple Inc.

[10] O'Hagan, A. and Stevens, J. W. (2001). Bayesian assessment of Sample size for clinical trails of cost-effctiveness. Med. Decision Making, 21, 219-230.

[11] Pezeshk, H. (2003). Bayesian techniques for Sample size determination in clinical trials: A short review. Statistical Method in Medical Research, 12, 6, 489-504.

[12] Pezeshk, H. and Gittins, J. C. (1999). Sample size Determination in Clinical Trials. Student, 3, 19-26.

[13] Pham-Gia, T. and Turkkan, N. (1992). Sample Size Determination in Bayesian Analysis (Disc: P399-404). The Statistician: Journal of the Institute of Statisticians, 41, 389-397.

[14] Raiffa, H. and Schalaifer, R. (1961). Applied statistical Decision Theory. Boston, Havard University Graduate School of Business Adminstration.

[15] Sahu, S. K. and Smith, T. M. F. (2006). A Bayesian method of sample size determination with practical applications. J. Roy. Statist. Soc. Ser. A, 169, 235–253.

[16] Saiful Islam, A. F. M. and Pettit, L. I. (2014). Bayesian sample size determination for the bounded linex loss function, Journal of Statistical Computation and Simulation, 84, 8, 1644-1653.

[17] Stallard, N. (1998). Sample size determination for phase II clinical trials based on Bayesian decision theory. Biometrics, 54, 279-294.

[18] Zellner, A. (1986) Bayesian estimation and prediction using asymmetric loss functions. Journal of the American Statistical Associations, 81, 394, 446-451.