Statistical Research and Training Center
مجله‌ی پژوهش‌های آماری ایران
©(2005) J. Stat. Res. Iran **2** 202-208
دوره‌ی ۱، شماره‌ی ۲، پاییز و زمستان ۱۳۸۳، صص ۱۰۱-۱۰۷
پژوهشکده آمار

# تعیین اندازه‌ی نمونه‌ی یک آزمایه با استفاده از روش لیندلی

حمید پزشک

دانشگاه تهران

**چکیده.** در این مقاله اندازه‌ی بهینه‌ی یک آزمایه‌ی تصادفی را برای حالتی که تابع مطلوبیت، تابعی از نسبت میانگین به واریانس تابع چگالی پسینِ پارامتر مورد علاقه است به دست می‌آوریم و از روش بیزی که رایفا و شلایفر (۱۹۶۱) به کار گرفتند و توسط لیندلی (۱۹۹۷) و اوهاگان و استیونس (۲۰۰۱) توسعه یافت، استفاده می‌کنیم. فرض‌هایی که در نظر خواهیم گرفت از یک دیدگاه، مسئله را بسیار ساده می‌کند و شاید با آنچه در عمل رخ می‌دهد، تفاوت به‌وجود آورد اما هدف این است که نشان دهیم مسئله در یک حالت خاص دارای جواب صریح و قابل تعمیم به حالت‌های پیچیده‌تر است.

**واژگان کلیدی.** اندازه‌ی نمونه؛ رهیافت بیزی؛ تابع مطلوبیت؛ هزینه؛ سود خالص؛ روش MEU.

## ۱ مقدمه

از زمانی که گروندی و همکاران (۱۹۵۶) مسئله‌ی تعیین اندازه‌ی نمونه‌ی یک آزمایه‌ی تصادفی را با به‌کارگیری روش‌های نظریه‌ی تصمیم بیزی معرفی کردند، نزدیک به نیم قرن می‌گذرد. این روش که در سال ۱۹۹۷ توسط لیندلی دوباره مطرح شد، توجه بسیاری از پژوهشگران را به خود جلب کرده است. استالارد (۱۹۹۸) با به‌کارگیری این روش در فاز دوم آزمایه‌های بالینی، اندازه‌ی نمونه را با ماکسیمم‌سازی تابع مطلوبیتی خاص به دست می‌آورد. پزشک و گیتینز (۱۹۹۹) مسئله‌ی تعیین اندازه‌ی نمونه را با استفاده از روش بیز برای

[ Downloaded from jsri.srtc.ac.ir on 2025-11-29 ]

[ DOI: 10.18869/acadpub.jsri.1.2.101 ]

یک آزمایه‌ی بالینی مطرح و آن را برای حالتی خاص حل می‌کنند. اوهاگان و استیونس (۲۰۰۱) یک تابع هدف شامل هزینه و سود انجام یک آزمایه‌ی تصادفی مطرح می‌کنند و با بهینه‌سازی آن تابع هدف، اندازه‌ی نمونه‌ی آزمایه‌ی تصادفی را به‌روش بیزی به دست می‌آورند. پزشک (۲۰۰۳) مروری از روش‌های مختلف بیزی برای تعیین اندازه‌ی نمونه‌ی آزمایه‌های تصادفی به دست می‌دهد.

در این مقاله با به‌کارگیری روش رایفا و شلایفر (۱۹۶۱) اندازه‌ی نمونه‌ی یک آزمایه‌ی تصادفی را با توجه به رهیافت کاملاً بیزی به دست می‌آوریم. فرض‌هایی را که به کار می‌گیریم، تا حدی از کلّیت مسئله در عمل می‌کاهند ولی هدف اصلی این است که نشان دهیم مسئله در حالت خاص دارای جواب صریح و قابل تعمیم به حالت‌های کلّی‌تر است.

فرض می‌کنیم داده‌ها از یک توزیع نرمال $\mathrm{N}(\theta, \sigma^2)$ که در آن $\sigma^2$ معلوم و $\theta$ نامعلوم است به دست آمده‌اند. برای $\theta$ یک چگالی پیشین در نظر می‌گیریم.

در بخش ۲، صورت مسئله‌ی تعیین اندازه‌ی نمونه را به‌طور رسمی مطرح می‌کنیم. نمادها و تابع هدف را نیز در این بخش معرفی می‌کنیم.

در بخش ۳ با معرفی یک تابع مطلوبیت خاص، مسئله‌ی تعیین اندازه‌ی نمونه را از دیدگاه بیز مطرح می‌کنیم و به بهینه‌سازی تابع هدف می‌پردازیم تا اندازه‌ی نمونه‌ی آزمایه‌ی تصادفی را بر اساس ماکسیمم‌سازی سود حاصل از آزمایه‌ی تصادفی به دست آوریم. سپس نقاط قوت و ضعف الگوی معرفی شده را مورد بررسی قرار می‌دهیم و یک مثال ارائه می‌کنیم.

## ۲ نمادها

فرض کنید $X_i$، $(i = 1, 2, \ldots)$، دنباله‌ای از متغیّرهای تصادفی مستقل نرمال $\mathrm{N}(\theta, \sigma^2)$ باشند که در آن $\sigma^2$ معلوم است. همچنین، فرض کنید $\theta$ دارای توزیع پیشین نرمال $\mathrm{N}(\mu, \tau^2)$ است. اگر $\overline{X}$ میانگین یک نمونه‌ی $n$تایی از $X_i$ها باشد آن‌گاه خواهیم داشت $\theta|\bar{x}_n \sim \mathrm{N}(\mu', \tau'^2)$ که در آن

$$\mu' = \frac{\sigma^2\mu + n\tau^2\bar{x}_n}{\sigma^2 + n\tau^2}, \qquad \tau'^2 = \frac{\sigma^2\tau^2}{\sigma^2 + n\tau^2}. \tag{1}$$

حال با استفاده از روش MEU (Maximization of Expected Utility) که توسط لیندلی (۱۹۹۷) ارائه شده است، تابع هدف مسئله را می‌نویسیم و اندازه‌ی نمونه را به دست می‌آوریم. در این روش، تابع مطلوبیت با $\mathrm{U}(n, \boldsymbol{x}, a, \theta)$ نشان داده می‌شود که با آن می‌توان سود نمونه‌ای به‌اندازه‌ی $n$ را در به دست آوردن نتیجه‌ی $\boldsymbol{x}$، وقتی تصمیم $a$ را می‌گیریم و مقدار پارامتر $\theta$ است، به دست آورد.



[ Downloaded from jsri.srtc.ac.ir on 2025-11-29 ]

[ DOI: 10.18869/acadpub.jsri.1.2.101 ]

اندازه‌ی نمونه‌ی $n$ در این روش به‌صورت زیر به دست می‌آید:

$$(۲) \qquad r(n) = \max_n \left\{ \int_x \max_a \left( \int_\theta u(n, \boldsymbol{x}, a, \theta) f^n(\theta|a, \boldsymbol{x})\ d\theta \right) f(\boldsymbol{x}|n)\ d\boldsymbol{x} \right\},$$

که در آن $f^n(\theta|a, \boldsymbol{x})$ تابع چگالی پسین و $f(\boldsymbol{x}|n)$ تابع چگالی پیشگو است. فرض می‌شود که تابع مطلوبیت به $\boldsymbol{x}$ بستگی ندارد و تابعی خطی به‌صورت زیر است:

$$u(n, \boldsymbol{x}, a, \theta) = u(a, \theta) - cn,$$

که در آن $c$ عددی ثابت است و $cn$ بیش‌تر به‌عنوان هزینه‌ی نمونه‌ای به‌اندازه‌ی $n$ در نظر گرفته می‌شود. بنا بر این، رابطه‌ی (۲) را می‌توان به‌صورت زیر ساده‌تر کرد:

$$(۳) \qquad r(n) = \max_n \left\{ \int_x \max_a \left( \int_\theta u(a, \theta) f^n(\theta|\boldsymbol{x})\ d\theta \right) f(\boldsymbol{x}|n)\ d\boldsymbol{x} - cn \right\}.$$

توجه کنید که انتخاب $a$ تأثیری بر $\theta$ به‌شرط $\boldsymbol{x}$ و $n$ ندارد، پس در رابطه‌ی (۳) به‌جای $f^n(\theta|a, \boldsymbol{x})$ نماد $f^n(\theta|\boldsymbol{x})$ را قرار داده‌ایم.

همان‌طور که توسط لیندلی (۱۹۹۷) اشاره شد، روش MEU از دیدگاه نظری روشی مناسب برای تعیین اندازه‌ی نمونه است، زیرا نخست آن‌که روشی منسجم است، دوم آن‌که یک الگوریتم خوش‌تعریف مشتمل بر یک سری امید ریاضی و ماکسیمم‌سازی برای آن موجود است، و سوم آن‌که بسیار انعطاف‌پذیر است زیرا طیفی گسترده از توابع مطلوبیت را می‌توان در آن به کار گرفت.

با به‌کارگیری روش MEU تابع هدف مسئله‌ی مورد نظر ما به‌صورت زیر خواهد بود:

$$r(n) = \int_{\bar{x}_n} \int_\theta u(\bar{x}_n, \theta) f^n(\theta|\bar{x}_n) f(\bar{x}_n)\ d\theta\ d\bar{x}_n - cn,$$

که در آن می‌خواهیم اندازه‌ی نمونه‌ی $n$ را به‌گونه‌ای بیابیم که تابع بالا ماکسیمم شود.

با استفاده از رابطه‌ی (۱) و توجه به این مطلب که $\overline{X}_n$ دارای چگالی $\mathrm{N}(\mu_۱, \frac{n\tau^۲+\sigma^۲}{n})$ است، داریم:

$$r(n) = \int_{\bar{x}_n} \int_\theta u(\bar{x}_n, \theta) \sqrt{\frac{n}{۲\pi(n\tau^۲ + \sigma^۲)}}\ \exp\{\frac{-n}{۲(n\tau^۲ + \sigma^۲)}(\bar{x}_n - \mu)^۲\} d\bar{x}_n - cn.$$

در بخش بعدی تابع مطلوبیت را به‌عنوان تابعی از میانگین چگالی پسین معرفی می‌کنیم.



[ Downloaded from jsri.srtc.ac.ir on 2025-11-29 ]

[ DOI: 10.18869/acadpub.jsri.1.2.101 ]

## ۳ تعیین اندازه‌ی نمونه

فرض کنید تابع مطلوبیت به‌صورت زیر باشد:

$$u(\bar{x}_n, \theta) = \left\{ k_1 \left( \frac{\mu' - \theta_0}{\tau'} \right) + k_2 \right\} b,$$

که در آن $k_1$ و $k_2$ و $\theta_0$ و $b$ اعداد ثابت‌اند و $k_1 > 0$ و $b > 0$.

یکی از مزایای به‌کارگرفتن چنین تابع مطلوبیتی این است که اگر $\theta_0$ میزان تأثیر یک نوع داروی خاص موجود در بازار باشد و $\theta$ میانگین تأثیر داروی جدیدی باشد که قرار است وارد بازار شود و به‌خاطر عدم قطعیتی که در مورد $\theta$ داریم، یک تابع چگالی پیشین برای آن در نظر گرفته‌ایم، آن‌گاه این تابع مطلوبیت می‌تواند به‌عنوان سود حاصل از مصرف بعدی داروی جدید در نظر گرفته شود. یعنی، چنان‌چه در اثر مصرف هر واحد دارو، $b$ واحد سود نصیب داروساز شود آن‌گاه $u(\bar{x}_n, \theta)$ حاصل‌ضرب تعداد کلّ مصرف‌کنندگان بعدی و سود هر واحد را نشان می‌دهد.

اگر $z_i = x_i - \theta_0$، $(i = 1, 2, \ldots)$ را در نظر بگیریم، آن‌گاه تابع مطلوبیت را می‌توانیم به‌شکل زیر ساده‌تر کنیم:

$$u(\bar{z}_n, \theta - \theta_0) = b \left\{ k_1 \left( \frac{\mu'}{\tau'} \right) + k_2 \right\}, \tag{۴}$$

که در آن $\mu'$ و $\tau'$ میانگین و انحراف معیار توزیع پسین $\delta = \theta - \theta_0$ هستند.

در این صورت تابع هدف می‌تواند به‌شکل زیر نوشته شود:

$$r(n) = \int_{\bar{z}_n} \int_{\theta} u(\bar{z}_n, \theta - \theta_0) f^n(\theta | \bar{z}_n) f(\bar{z}_n) \, d\theta \, d\bar{z}_n - cn. \tag{۵}$$

با جایگذاری رابطه‌ی (۴) در (۵) و محاسبه‌ی انتگرال داخلی، سرانجام، تابع هدف به‌شکل زیر ساده می‌شود:

$$r(n) = \int_{\bar{z}_n} \left\{ k_1 \left( \frac{\mu'}{\tau'} \right) + k_2 \right\} b f(\bar{z}_n) \, d\bar{z}_n - cn.$$

با جایگذاری مقادیر $\mu'$ و $\tau'$ برای $\delta = \theta - \theta_0$ خواهیم داشت:

$$r(n) = \int_{\bar{z}_n} k_1 b \left\{ \left( \frac{\sigma^2 \mu + n\tau^2 \bar{z}_n}{\sigma^2 + n\tau^2} \right) \times \left( \frac{1}{\sqrt{\frac{\sigma^2 \tau^2}{\sigma^2 + n\tau^2}}} \right) \right\} f(\bar{z}_n) \, d\bar{z}_n$$
$$+ \int_{\bar{z}_n} k_2 b f(\bar{z}_n) d\bar{z}_n - cn.$$



[ Downloaded from jsri.srtc.ac.ir on 2025-11-29 ]

[ DOI: 10.18869/acadpub.jsri.1.2.101 ]

از این‌جا خواهیم دید:

$$\frac{r(n)}{k_1 b} - \frac{k_2}{k_1 b}\left(\frac{\mu}{\sigma}\right) = \left\{\left(\frac{\tau^2}{\sigma}\right)^{-2} + n\right\}^{-\frac{1}{2}}\frac{\mu}{\sigma} - \frac{c}{k_1 b}n.$$

با تغییر متغیّر $R(n) = \frac{r(n)}{k_1 b} - \frac{k_2}{k_1 b}\left(\frac{\mu}{\sigma}\right)$ خواهیم داشت:

$$R(n) = (T^{-2} + n)^{\frac{1}{2}} D - Cn,$$

که در آن $C = \frac{c}{k_1 b}$، $D = \frac{\mu}{\sigma}$ و $T = \frac{\tau}{\sigma}$.

برای به دست آوردن اندازه‌ی نمونه‌ی بهینه فرض می‌کنیم $R(n)$ تابعی پیوسته از $n$ است و قرار می‌دهیم $\frac{dR(n)}{dn} = 0$ :

$$\frac{dR(n)}{dn} = \frac{1}{2}(T^{-2} + n)^{-\frac{1}{2}} D - C = 0.$$

به‌سادگی اثبات می‌شود که آن مقدار از $n$ که تابع هدف بالا را ماکسیمم می‌کند، برابر است با:

$$n^* = \left[\left(\frac{D}{2C}\right)^2 - T^{-2}\right] + 1, \qquad (6)$$

که در آن $[x]$ نشانه‌ی جزء صحیح $x$ است.

توجه به این نکته ضروری است که اگر $C > \frac{1}{2}DT$ آن‌گاه $n^* = 0$ و این بدان معنی است که هزینه‌ی انجام آزمایه بیش از سودی است که از آن به دست خواهد آمد.

## ۳٫۱ پارامترهایی که در تعیین اندازه‌ی نمونه مؤثرند

از رابطه‌ی (۶) متوجه می‌شویم که مقدار بهینه‌ی اندازه‌ی نمونه فقط تابعی از پارامترهای $C$، $T$ و $D$ است که می‌توان آن‌ها را به شکل زیر تعبیر کرد:

$C$: نسبت هزینه به سود،

$T$: نوعی اندازه‌گیری از جهل پیشین ما درباره‌ی $E_{prior}(\delta) = \mu - \theta_0$،

$D$: نوعی اندازه‌گیری از امید ریاضی پیشین ما از بیش‌تر بودن $\theta$ از $\theta_0$.

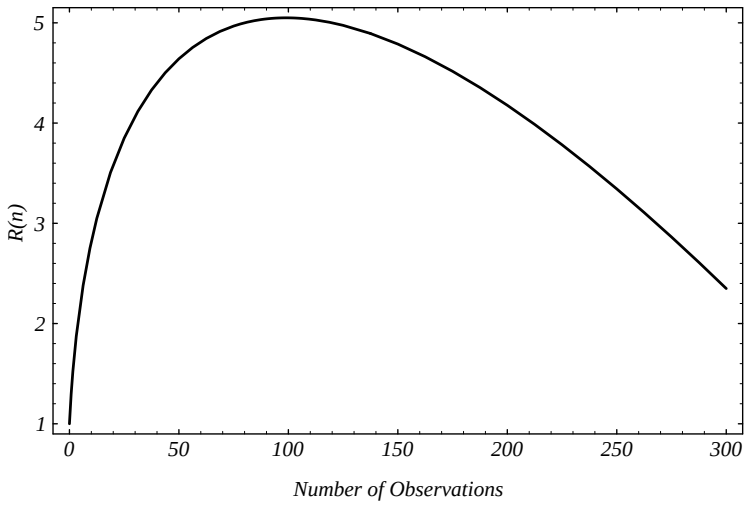
**مثال.** در آزمایه‌ای برای استنباط درباره‌ی $\delta$ (که می‌تواند میزان متوسط برتری تیمار جدید نسبت به تیمار



[ Downloaded from jsri.srtc.ac.ir on 2025-11-29 ]

[ DOI: 10.18869/acadpub.jsri.1.2.101 ]

قدیمی در یک آزمایه‌ی بالینی در نظر گرفته شود) نسبت هزینه به سود $C = ۰٫۰۵$، $T = ۱$ و $D = ۱$ فرض شده‌اند. تابع هدف $R(n)$ را نسبت به $n$ رسم و $n^*$ را بر اساس رابطه‌ی (۶) محاسبه می‌کنیم.




نمودار ۱. تابع هدف یا امید ریاضی سود خالص برای $D = ۱$، $T = ۱$ و $C = ۰٫۰۵$.

اندازه‌ی بهینه برای آزمایه‌ای که تابع هدف آن در نمودار ۱ رسم شده برابر است با $n^* = ۹۹$. محاسبات نشان می‌دهد که $n^*$ تابعی نزولی از $C$ ولی تابعی صعودی از هر دوی $T$ و $D$ است. این‌که $n^*$ باید تابعی نزولی از نسبت هزینه به سود باشد و همچنین صعودی بودن $n^*$ به‌عنوان تابعی از $T$، منطقی است اما چرا وقتی ما انتظار داریم $\theta$ از $\theta_۰$ بیش‌تر باشد نیاز به اندازه‌ی نمونه‌ی بزرگ‌تر داریم؟ یعنی این‌که $n^*$ تابعی صعودی از $D$ است که غیرمنطقی به‌نظر می‌رسد. این ضعف در واقع به‌علت انتخاب ما از تابع مطلوبیت به‌وجود آمده است که آن را تابعی خطی از $\mu'$ یعنی امید ریاضی چگالی پسین $\delta = \theta - \theta_۰$ در نظر گرفتیم و اجازه دادیم تابع $k_۱\left(\frac{\mu'}{\tau^۲}\right) + k_۲$ تا بی‌نهایت تغییر کند. یک اصلاح در تابع مطلوبیت که آن را از بالا کراندار کند از این مشکل جلوگیری خواهد کرد.

## مرجع‌ها

Grundy, P.M.; Healy, M.J.R.; Rees, D.H. (1956). Economic Choice of the Amount of Experimentation. *J. R. Statist. Soc.* A **18**, 32-48.



[ Downloaded from jsri.srtc.ac.ir on 2025-11-29 ]

[ DOI: 10.18869/acadpub.jsri.1.2.101 ]

Lindley, D.V. (1997). The Choice of Sample Size. *The Statistician* **46**, 129-166.

ÓHagan, A.; Stevens, J.W. (2001). A Framework for Cost-Effectiveness Analysis of Clinical Trial Data. *Health Econ.* **10**, 302-315.

Pezeshk, H. (2003). Bayesian Techniques for Sample Size Determination Clinical Trials: A Short Review. *Stat. Methods Med. Res.* **12**, 489-504.

Pezeshk, H.; Gittins, J. (1999). Sample Size Determination in Clinical Trials. *Student* **3**, 19-26.

Raiffa, H.; Schlaifer, R. (1961). *Appl. Statist. Decis. Theory.* Cambridge, MA, USA. Harvard University Graduate School of Business Administration, Division of Research.

Stallard, N. (1998). Sample size Determination for Phase II Clinical Trials Based on Bayesian Decision Theory. *Biometrics* **54**, 279-294.

دریافت: ۴ خرداد ۱۳۸۳
آخرین اصلاح: ۹ مهر ۱۳۸۴

**حمید پزشک**
قطب زیست‌ریاضی (بیومتمتیک)،
دانشکده‌ی ریاضی، آمار و علوم کامپیوتر،
پردیس علوم، دانشگاه تهران،
خیابان انقلاب اسلامی،
تهران، ایران.
پیام‌نگار: *pezeshk@khayam.ut.ac.ir*



[ Downloaded from jsri.srtc.ac.ir on 2025-11-29 ]

[ DOI: 10.18869/acadpub.jsri.1.2.101 ]

[ Downloaded from jsri.srtc.ac.ir on 2025-11-29 ]

[ DOI: 10.18869/acadpub.jsri.1.2.101 ]