

# بررسی عوامل مؤثر بر زمان بقای بیماران مبتلا به لوسمی حاد فوت شده و بر آورد میانگین زمان بقا توسط الگوریتم Expectation & Maximation و روش شبیه سازی مونت کارلویی زنجیره‌های مارکف

دکتر محمد بهرامی\*، دکتر محمدرضا مشکانی\*\*، دکتر مژگان اعلم صمیمی\*\*\*

\* عضو هیأت علمی، گروه آمار دانشگاه اصفهان

\*\* استاد آمار، گروه آمار دانشگاه شهید بهشتی

\*\*\* فوق تخصص بیماریهای خون و سرطان، عضو هیأت علمی دانشگاه علوم پزشکی اصفهان

تاریخ دریافت: ۱۳۸۵/۱۰/۲۷

تاریخ پذیرش: ۱۳۸۵/۱۲/۱۱

## چکیده

لوسمی یکی از انواع بیماری‌های بدخیم دستگاه خونساز بدن است که در مدت زمان بسیار کوتاهی باعث مرگ و میر می‌شود؛ در این مقاله، عوامل مؤثر بر زمان بقای بیماران مبتلا به لوسمی حاد لنفوتییدی (ALL)، مورد بررسی قرار گرفته است تا برای درک ارتباط میان زمان بقای این بیماران با عوامل مؤثر Expectation & Maximation، مدل رگرسیون پیش‌بینی کننده مناسبی به دست آید.

اطلاعات ۵۲ بیمار لوسمی فوت شده در بیمارستان سیدالشهدای (ع) اصفهان بررسی شد. مدل رگرسیون پیش‌بینی کننده مد نظر شامل متغیرهای هموگلوبین، سلول‌های نابالغ و سن بیمار بود. به دلیل توزیع آمیخته داده‌های موجود، یک مدل آمیخته برای زمان بقای بیماران در نظر گرفته، با استفاده از الگوریتم EM، برآورد بیشینه درست نمایی میانگین بقا را محاسبه کردیم. همچنین با به کارگیری روش شبیه سازی مونت کارلویی زنجیره‌های مارکف (MCMC) برآورد بیزی میانگین بقا را به دست آوردیم.

با استفاده از روش‌های آماری فوق، تابع بقایی را به دست آوردیم که با استفاده از آن می‌توان مدت زمان بقای بیماران را بر حسب روز (به‌عنوان پیش‌آگهی) پیش‌بینی کرده در مورد درمان بیمار تصمیم گرفت. به نظر می‌رسد می‌توان با مطالعات با حجم بیشتر و استفاده از روش‌ها و آزمون‌های آماری ذکر شده در این مقاله، ارتباطی بین یافته‌های بالینی و آزمایشگاهی و میزان بقا پیدا کرد. لذا می‌توان از این مدل برای سایر بیماری‌ها نیز جهت تعیین پیش‌آگهی استفاده نمود.

برآورد بیشینه درست نمایی، برآورد بیزی، دو نمایی، رگرسیون، لوسمی، مدل آمیخته، میانگین بقا

مقدمه:

روش‌ها:

یافته‌ها:

نتیجه‌گیری:

واژگان کلیدی:

تعداد صفحات: ۹

تعداد جدول‌ها: ۵

تعداد نمودارها: ۲

تعداد منابع: ۲۳

دکتر محمد بهرامی، گروه آمار دانشگاه اصفهان

آدرس نویسنده مسئول:

E-mail: bahrami@sci.ui.ac.ir

## مقدمه

لوسمی حاد لنفوبلاستیک (Acute Lymphoblastic Leukemia) یکی از بیماری‌هایی است که با افزایش سلول‌های نارس در مغز استخوان، خون و سایر بافت‌ها مشخص می‌شود و به‌همین دلیل می‌تواند در خونسازی طبیعی و دستگاه ایمنی مشکل ایجاد کند. در بعضی از بیماران درمان قطعی انجام می‌گیرد اما در برخی دیگر، با وجود انجام پیوند مغز استخوان، متأسفانه بیماری مهار نشده، باعث مرگ می‌شود (۱). موفقیت درمان لوسمی حاد در کودکان حدود ۸۰٪ گزارش شده است (۱) در حالی که این میزان در بزرگسالان، حتی با وجود استفاده از پیوند مغز استخوان به ندرت به ۴۰٪ می‌رسد (۹-۲). در بدو تشخیص می‌توان از برخی علائم بالینی و پاراکلینیکی مانند سن و تعداد گلبول‌های سفید به‌عنوان عوامل تعیین‌کننده پیش‌آگهی بهره برد؛ به‌طوری که در بزرگسالان با افزایش سن و افزایش تعداد گلبول‌های سفید پیش‌آگهی بدتر می‌شود. با این وجود، هیچ راهنمای دقیقی برای تعیین وضعیت بزرگسالان با ریسک بالا وجود ندارد (۶-۳، ۵) و تلاش برای جداکردن افراد دارای پیش‌آگهی خوب از افراد دارای پیش‌آگهی بد ادامه دارد. به همین دلیل تصمیم گرفتیم تا در بیماران مبتلا به لوسمی حاد که در بیمارستان سید الشهداء (ع) اصفهان تحت درمان قرار گرفته و فوت شده بودند، ارتباط عواملی مانند سن، جنس، گلبول‌های سفید، پلاکت‌ها، لنفوسیت‌ها و سلول‌های نابالغ را با زمان بقای آن‌ها مورد مطالعه قرار دهیم تا با توجه به تطابق یافته‌های به‌دست آمده در مدل پیش‌بینی‌کننده و برآورد میانگین بقا با دو روش کلاسیک و بیزی بتوان در سایر بیماران نیز از این مدل

استفاده و از ابتدا پیش‌آگهی را در آن‌ها مشخص کرد. قابل ذکر است که این موضوع در منابع خارجی نیز مورد بررسی قرار گرفته است اما، با توجه به این که به نظر می‌رسد طبیعت بیماری‌های مختلف در بیماران ما با بیماران مطالعه‌شده در آن منابع متفاوت باشد، مطالعه این عوامل در بیماران جوامع مختلف، برای پیش‌بینی جواب به درمان مفیدتر به نظر می‌رسد.

## روش‌ها

برای بررسی عوامل مؤثر بر زمان بقای بیماران مبتلا به لوسمی حاد، نمونه‌ای به حجم پنجاه و دو نفر از بیماران لوسمی حاد فوت شده در بیمارستان امید اصفهان از سال ۱۳۷۳ تا سال ۱۳۸۲ جمع‌آوری و اطلاعات لازم از پرونده‌های این بیماران استخراج گردید. این اطلاعات مربوط به اولین آزمایش تشخیصی انجام شده قبل از هرگونه اقدام درمانی بود. بنابراین آزمایشات فرمول شمارش خون (CBC) مربوط به این بیماران با سنین مختلف و دارای شرایط تقریباً همگون درمانی به منظور بررسی عوامل جنس، سن، گلبول‌های سفید (WBC)، هموگلوبین (HGB)، پلاکت‌ها (PLT)، لنفوسیت‌ها (LYMP) و سلول‌های نابالغ (LUC) بازبینی گردید. دامنه طبیعی هر یک از فاکتورهای خونی مورد نظر در جدول ۱ آورده شده است.

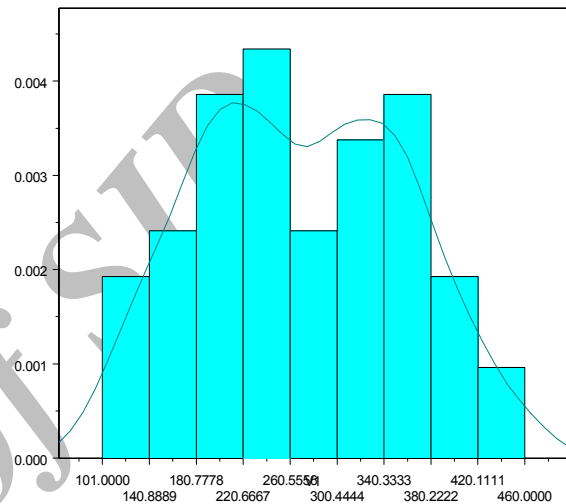
جدول ۱. دامنه نرمال فاکتورهای خونی مورد نیاز

۱۵۰-۴۵۰×۱۰	PLT	۴/۳-۱۰/۸×۱۰	WBC
۰/۱۶-۰/۴۵×۱۰/ml	LYMP	۱۸-۱۳gr/dl	HGB: مؤنث: افراد
۰-۰/۰۴×۱۰/ml	LUC	۱۲-۱۶gr/dl	HGB: مذکر: افراد

لازم به ذکر است که تشخیص بیماری لوسمی حاد توسط اسپیراسیون مغز استخوان انجام گرفته بود.

## یافته‌ها

برای پیش بینی زمان بقای بیماران ALL فوت شده، ابتدا نرمال بودن متغیر زمان بقای بیماران یعنی متغیر پاسخ Y مورد بررسی قرار گرفت. برای این منظور ابتدا هیستوگرام و منحنی فراوانی نسبی این داده‌ها را رسم کردیم.



شکل ۱. فراوانی نسبی زمان بقای بیماران لوسمی حاد بر حسب روز

گام بعدی، بررسی این موضوع بود که متغیر پاسخ با کدامیک از متغیرهای مستقل سن، WBC، PLT، LYMP، HGB، LUC و جنس رابطه خطی دارد. برای این منظور ضریب همبستگی پیرسن (Pearson Correlation Coefficient) را میان متغیر پاسخ و تمام متغیرهای مستقل به دست آوردیم. مقادیر عددی به دست آمده حاکی از آن بود که بین متغیر پاسخ Y و دو متغیر HGB و LUC یک رابطه خطی به نسبت قوی وجود دارد. زیرا:

$$\text{Correlation (Y, HGB)} = 0.745; (p=0.000)$$

$$\text{Correlation (Y, LUC)} = -0.61; (p=0.000)$$

لازم به ذکر است، چون متغیر Gender یک متغیر کیفی می‌باشد، برای بیان ارتباط میان این متغیر و متغیر Y از ضریب همبستگی اسپیرمن استفاده شده است که مقدار به دست آمده آن برابر است با:

$$\text{Spearman Correlation (Y, Gender)} = -0.022; (P=0.878)$$

که نشان می‌دهد این متغیر ارتباط خطی با متغیر پاسخ Y ندارد.

با مشاهده دقیق داده‌ها به نظر رسید که میان سن و متغیر پاسخ Y نیز نوعی ارتباط وجود داشته باشد. به این منظور سن بیماران را به سه گروه تقسیم کردیم: گروه اول ۰-۲۴، گروه دوم ۲۵-۴۹ و گروه سوم ۵۰ به بالا. علت این گروه‌بندی، بیشتر بودن زمان بقا در گروه اول و سوم نسبت به گروه دوم است. این گروه‌بندی نشان داد که یک رابطه غیر خطی، مانند یک سهمی، میان زمان بقا و متغیر Age برقرار است (شکل ۲). بنابراین، برای برقراری رابطه خطی میان متغیر پاسخ Y و Age، توابع مختلفی از سن مانند تابع‌های قدر مطلق و درجه دوم را بررسی کردیم که تنها یک تبدیل مؤثر شناخته شد؛  $\text{Age}_1 = (\text{Age} - 42)^2$  که در آن عدد ۴۲ متناظر با کم‌ترین زمان بقا و عدد ۴۴۱ واریانس

نمودار ۱ نشان می‌دهد که متغیر پاسخ Y تقریباً دارای توزیع نرمال است. برای اطمینان بیشتر از نرمال بودن داده‌ها، با آزمون غیرپارامتری کلموگروف-اسمیرنوف (Kolmogrov-Smirnov Test) فرض نرمال بودن داده‌ها را بررسی کردیم. مقدار آماری آزمون عدد ۰/۰۸۲۲ به دست آمد، که کوچکتر از مقدار آن در جدول (۰/۱۶ برای سطح ۰/۰۵) است؛ بنا براین فرض نرمال بودن داده‌ها رد نمی‌شود. همچنین برای بررسی تصادفی بودن داده‌ها از آزمون ناپارامتری Runs-Test استفاده کردیم. در این حالت نیز فرض تصادفی بودن داده‌ها با میزان ۰/۰۵ پذیرفته شد. بنابراین همه این آزمون‌ها مشخص شد که متغیر پاسخ Y یک نمونه تصادفی مستقل با توزیع نرمال است.

برای این مدل، مقدار  $R^2$  (ضریب تعیین اصلاح شده) برابر با  $0.7282$  به دست آمد که تا حد زیادی حاکی از درست بودن آن است. جدول آنالیز واریانس نیز صحت مدل رگرسیونی به دست آمده را تأیید می‌کند.

جدول ۳. جدول آنالیز واریانس مدل رگرسیونی زمان بقای بیماران

لوسمی حاد						
درجه آزادی	مجموع مربعات	مینگین توان دوم خطا	مقدار $F$ احتمال (F)			
۱	۲۲۰۰۸۲/۰	۲۲۰۰۸۲/۰	۹۷/۹۰۱۵۴	۰/۰۰۰۰	HGB	
۱	۴۷۸۹۳/۰	۴۷۸۹۳/۰	۲۱/۳۰۴۷۹	۰/۰۰۰۰۲۹	LUC	
۱	۲۱۰۶۵/۸	۲۱۰۶۵/۸	۹/۳۷۰۹۱	۰/۰۰۳۶	Age <sub>1</sub>	
۴۸	۱۰۷۹۰۳/۷	۰,۲۲۴۸			باقیمانده‌ها	

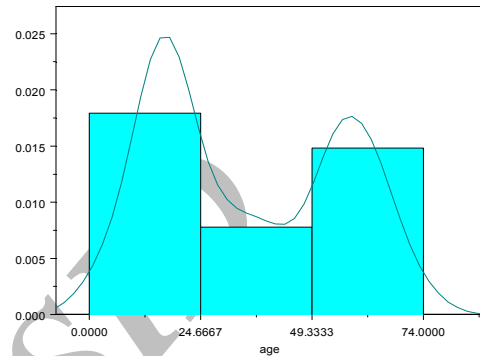
بنا بر این مدل رگرسیونی (پیش‌بینی کننده) مورد نظر به صورت زیر به دست آمد:

$$\hat{Y} = 81.0353 + 22.5674 * HGB - 6.3590 * LUC + 24.4034 * Age$$

به عنوان مثال اگر یک بیمار مبتلا به لوسمی حاد در زمان تشخیص بیماری ۵۵ سال داشته باشد و در اولین آزمایش‌ها، میزان HGB خون او  $7/6$  و میزان LUC خون او  $8/1$  باشد، در این صورت پیش‌بینی می‌شود، مدت ۲۱۰ روز زنده بماند. در مدل ۱ ضریب تعیین مدل:  $R^2 = 0.7282$  به دست آمد و حاکی از آن بود که برازش مدل تا حد زیادی درست است. با توجه به مقدار  $p$ -value در جدول آنالیز واریانس مدل نیز درمی‌یابیم که فرض‌های  $H_0: \beta_1 = 0$ ،  $H_0: \beta_2 = 0$  و  $H_0: \beta_3 = 0$  با میزان ۵٪ رد می‌شوند. یعنی سه متغیر HGB، LUC، و Age<sub>1</sub> باید در مدل رگرسیونی مورد نظر باشند.

فراوانی بیماری لوسمی در مردان کمی بیشتر از زنان گزارش شده است و به طور متوسط زمان بقای مردان مبتلا به لوسمی از زنان کم‌تر است (۱). بدین منظور بعد از بررسی این موضوع که داده‌های زمان بقای

داده‌های سن است. با این تبدیل، ضریب همبستگی پیرسن میان Y و متغیر جدید سن که آن را با Age<sub>1</sub> نشان داده ایم، عدد ۴۴ به دست آمد ( $P = 0/005$ ). در نتیجه، به نظر رسید متغیرهایی که باید در مدل وجود داشته باشند LUC، HGB، و Age<sub>1</sub> هستند.



شکل ۲. هیستوگرام فراوانی نسبی مربوط به سن (سال) بیماران فوت شده در اثر لوسمی حاد

برای سایر متغیرها نیز تبدیل‌های مختلفی را بررسی کردیم؛ اما، هیچ‌یک مفید واقع نشد و نتوانست یک رابطه خطی با Y برقرار کند. با توجه به تمام معیارهای موجود، برای انتخاب مدل مناسب و به کارگیری روش گام به گام پیشرو (Stepwise Forward Method)، بهترین مدلی که برای پیش‌بینی زمان بقای بیماران لوسمی فوت شده به دست آمد، مدلی شامل سه متغیر HGB، LUC، و Age<sub>1</sub> بود که ضرایب رگرسیونی آن با استفاده از نرم‌افزار S-PLUS به دست آمد (جدول ۲).

جدول ۲. مقادیر به دست آمده ضرایب رگرسیونی مدل برازش داده شده برای زمان بقای بیماران مبتلا به لوسمی حاد

ضرایب	مقدار ضرایب	خطای استاندارد	مقدار -t	احتمال (> t )
ثابت	۸۱/۰۳۵۳	۳۱/۹۸۴۷	۲/۵۳۳۶	۰/۰۱۴۶
HGB	۲۲/۵۶۷۴	۳/۳۸۵۲	۶/۶۶۶۵	۰/۰۰۰۰
LUC	-۶/۳۵۹۰	۱/۴۰۴۷	-۴/۵۲۶۹	۰/۰۰۰۰
Age <sub>1</sub>	۲۴/۴۰۳۴	۷/۹۷۱۸	۳/۰۶۱۲	۰/۰۰۳۶

بقای ۵۲ بیمار لوسمی حاد فوت شده در بیمارستان سید الشهداء (ع) اصفهان به دست می‌آوردیم. از آنجایی که به دست آوردن برآوردهای مربوط به پارامترهای موجود در مدل ۲ بسیار دشوار بود، از الگوریتم EM استفاده کردیم (۱۲-۱۱). این الگوریتم از دو کلمه Expectation (امید ریاضی) و Maximization (بیشینه کردن) تشکیل شده است و در واقع بهترین روش عددی موجود برای برآورد پارامترها در مدل‌های آمیخته می‌باشد. در این روش ابتدا امید ریاضی شرطی متغیرهایی را که به آن‌ها متغیرهای پنهان (Latent Variables) گفته می‌شود به دست آورده، سپس با شروع از یک مقدار اولیه برای هر پارامتر، تابع درست نمایی را بیشینه می‌کنیم. برنامه رایانه‌ای روش EM را در نرم‌افزار Matlab نوشتیم و برآورد پارامترها را به دست آوردیم (جدول ۴).

جدول ۴: برآوردهای بیشینه درست نمایی پارامترهای مدل

پارامتر	مقدار برآورد	خطای استاندارد
$\pi$	۰/۴۶۴۰	۰/۰۰۱۳۶۱
$\mu_1$	۱۹۵/۵۰۷۴	۰/۱۹۲۶۵
$\mu_2$	۳۳۴/۶۵۹۶	۰/۲۳۳۸۱
$\sigma_1^2$	۲۳۰۰/۴۴۴۱	۷/۶۴۴۱۳
$\sigma_2^2$	۳۲۶۵/۱۹۰۵	۸/۶۱۹۵۹

با توجه به برآوردهای بیشینه درست نمایی پارامترها، اکنون می‌توانستیم برآورد بیشینه درست نمایی تابع بقا (Survival Function) را به دست آوریم (۱۵)؛ مدل ۳ نشانگر روش این محاسبه است و در آن تابع  $\Phi$ ، تابع توزیع نرمال استاندارد می‌باشد.

مردان و زنان هر دو نرمال می‌باشند، آزمون  $t$  برای تساوی میانگین زمان بقای این دو گروه انجام شد و در سطح ۰/۰۵ اختلاف معنی‌داری بین میانگین‌های زمان بقای مردان و زنان مشاهده نگردید. در پایان اثرات متقابل سه متغیر موجود در مدل نیز بررسی گردید اما، هیچ‌کدام مورد پذیرش واقع نشد. با مشاهده دقیق‌تر هیستوگرام و منحنی فراوانی نسبی مربوط به زمان بقای بیماران مبتلا به لوسمی (شکل ۱)، دریافتیم که با توجه به دو نمایی (Bimodal) بودن داده‌ها، توزیع زمان بقا در بیماران متفاوت بوده، می‌تواند دارای چگالی آمیخته (Mixture Density) باشد. پیش‌تر نشان دادیم توزیع طول عمر این بیماران تقریباً نرمال است، بنابراین می‌توانستیم نتیجه بگیریم که این چگالی آمیخته در واقع آمیخته‌ای از دو چگالی نرمال با میانگین‌ها و واریانس‌های متفاوت است؛ آزمون نیکویی برازش (Goodness of Fit Test) صحت مدل آمیخته از دو توزیع نرمال را تأیید می‌کند (۱۰).

مدل آمیخته زیر، یک مدل آمیخته نرمال با دو مؤلفه با میانگین‌های  $\mu_1$  و  $\mu_2$  و واریانس‌های  $\sigma_1^2, \sigma_2^2$  می‌باشد.

مدل ۲

$$f(x) = \pi \cdot \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2\sigma_1^2}(x-\mu_1)^2} + (1-\pi) \cdot \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{1}{2\sigma_2^2}(x-\mu_2)^2} \quad x \in \mathbf{R}$$

در مدل فوق،  $\pi$  را نسبت آمیختگی در مدل نامیدیم. مطالب نظری مربوط به چگونگی رفتار با این گونه مدل‌ها و انجام استنباط آماری درباره پارامترها را می‌توان در کتب آمار ملاحظه کرد (۱۰). در این جا تنها کاربرد آن نظریه را در قالب یک مسأله پزشکی آورده‌ایم. اکنون باید برآوردهای بیشینه درست نمایی برای مدل آمیخته (مدل ۲) را براساس داده‌های زمان

## مدل ۳

$$\hat{S}(t) = 0.464 * \left[ 1 - \Phi\left(\frac{t-195.51}{47.96}\right) \right] + 0.536 * \left[ 1 - \Phi\left(\frac{t-334.66}{57.14}\right) \right]$$

به عنوان مثال، احتمال این که یک بیمار مبتلا به لوسمی حاد بعد از تشخیص بیش از ۳۵۰ روز زنده باشد، چنین محاسبه می شود:  $\hat{S}(350) = 0.24$ .

از طرفی، اکنون می توانستیم برای تأیید مقادیر برآورد پارامترهای به دست آمده در جدول ۴ با انتخاب توزیع پیشین مناسب (۱۴) برای پارامترهای موجود در مدل ۲، با استفاده از روش شبیه‌سازی مونت کارلو-وی زنجیره‌های مارکف (MCMC: Monte Carlo Markov Chains Simulation)، برنامه‌ای را در نرم‌افزار WINBUGS که یک نرم‌افزار تخصصی برای مدل‌های بیزی است، بنویسیم. برآوردهای بیزی (۱۸-۹، ۱۴) این پارامترها را در جدول ۵ نشان داده‌ایم. در روش شبیه‌سازی مونت کارلویی زنجیره‌های مارکف با انتخاب توزیع پیشین مناسب برای پارامترهای مجهول در توزیع آمیخته، ابتدا داده‌هایی از این توزیع را شبیه‌سازی کرده، سپس با استفاده از اطلاعات موجود در این داده‌ها توزیع پسین مناسب را برای پارامترهای توزیع آمیخته به دست آوردیم و از میانگین این توزیع پسین، در برآورد کردن پارامترهای مجهول استفاده کردیم.

جدول ۵: خروجی برنامه برآورد بیزی پارامترهای آمیخته

داده‌های زمان بقای بیماران لوسمی در نرم افزار WINBUGS

پارامتر میانگین	انحراف استاندارد	خطای MC	٪۲/۵	میان	٪۹۷/۵	شروع نمونه
$\pi$	۰/۰۰۰۰۷	۰/۰۰۰۰۷	۰/۰۰۰۰۷	۰/۰۰۰۰۷	۰/۰۰۰۰۷	۲۱۰۰۰
$\mu_1$	۴/۲۸۷	۰/۰۳۵۱۹	۱۹۲/۶	۲۰۰/۹	۲۰۹/۲	۲۱۰۰۰
$\mu_2$	۴/۲۹۸	۰/۰۳۵۱۷	۳۴۱/۰	۳۴۹/۴	۳۵۷/۹	۲۱۰۰۰
$\sigma_1^2$	۱۳/۹۲۰	۱۶/۵۵	۱۵۰۸/۰	۲۹۸۴/۰	۲۹۸۶۹/۰	۲۱۰۰۰
$\sigma_2^2$	۱۳/۲۶۰	۱۴/۴۴	۱۴۱۰/۰	۲۷۶۹/۰	۲۷۶۴/۰	۲۱۰۰۰

مقادیر عددی برآوردهای بیزی به دست آمده در جدول ۵ درستی مقادیر به دست آمده برآوردهای پیشینه درست‌نمایی در جدول ۴ را تأیید می‌کند. به عنوان مثال مقدار برآورد عددی پارامتر  $\pi$  در جدول ۴ برابر با ۰/۴۶۴ و در جدول ۵ که مقدار عددی برآورد همین پارامتر که با  $\pi$  نشان داده شده است، برابر با ۰/۴۳۸۳ می باشد. در واقع نتیجه به دست آمده از دو روش آمار کلاسیک و بیزی جواب‌های بسیار نزدیک بوده است.

## بحث

یکی از مسائل مهم در بیماران مبتلا به لوسمی حاد تعیین پیش آگهی بیمار است؛ چرا که در اصل، این پیش آگهی تخمین زده شده می‌باشد که تصمیم‌گیری‌های درمانی بعدی را تعیین می‌کند (۳، ۶). با توجه به وارد شدن پیوند مغز استخوان در عرصه درمان و مرتالتی و مریدیتی این روش درمانی، باید ابتدا نسبت سود و زیان پیوند مغز استخوان در مورد بیماران سنجیده و سپس در مورد انجام آن تصمیم‌گیری شود (۲۰-۱۹، ۸)؛ بدیهی است این امر بدون تخمین پیش آگهی ممکن نخواهد بود. در حال حاضر در کشورهای توسعه یافته از سیتوژنتیک و روش‌های ملکولی برای تعیین پیش آگهی استفاده می‌شود؛ در حالی که، در دیگر کشورها امکان انجام این آزمایشات در سطح وسیع وجود ندارد. لذا تعیین تأثیر دقیق عوامل ابتدایی‌تر مانند سن، تعداد بلاست و شمارش اولیه بر پیش آگهی، از اهمیت بیشتری برخوردار می‌شود (۲۱) و اگر بتوان ارتباط آماری بین این عوامل پیدا کرد، می‌توان در مورد تصمیم‌گیری‌های بعدی درمانی، دقیق‌تر اظهار نظر نمود. در مطالعه حاضر سن اکثر بیماران بین ۲۵ تا ۴۹ بود اما طبق

ارتباط این پارامترها با زمان بقا با آزمون‌های آماری قابل بررسی بوده است، توصیه می‌شود این مدل در تمام بیماران مبتلا به ALL به کار برده شود.

همچنین ۴۶٪ این بیماران را افراد بین ۲۵ تا ۴۹ سال تشکیل می‌دادند و زمان بقای این دسته از بیماران به طور متوسط برابر با ۱۹۵ روز بود. سایر بیماران در سنین دیگر بوده و زمان بقای این دسته به طور متوسط ۳۳۵ روز بود.

نکته قابل توجه این است که اگرچه متغیرهای دیگری نیز می‌توانند بر زمان بقای بیماران مبتلا به لوسمی مؤثر باشند اما، داده‌های ما براساس اطلاعات موجود در پرونده‌های این بیماران بوده است و بنابراین تحلیل آماری انجام شده نیز براساس متغیرهای موجود می‌باشد.

**نتیجه‌گیری:** به نظر می‌رسد با انجام مطالعات با حجم نمونه بیشتر و استفاده از روش‌ها و آزمون‌های آماری ذکر شده در این مقاله، بتوان ارتباطی بین یافته‌های بالینی و آزمایشگاهی و میزان بقا پیدا کرد که در این صورت می‌توان از این مدل برای تعیین پیش آگهی سایر بیماری‌ها هم استفاده نمود.

منابع خارجی با افزایش سن شیوع ALL نیز افزایش می‌یابد (۲۱)؛ یعنی انتظار می‌رود که ALL در رده سنی ۴۹ به بعد از این تعداد بیشتر باشد. علت این موضوع ممکن است کوچک بودن حجم نمونه باشد و بنابراین نمی‌توان این یافته‌ها را تعمیم داد. از طرفی با افزایش سن، پیش آگهی و نیز زمان بقا در افراد مبتلا به ALL کاهش می‌یابد (۲۱) که در بیماران نمونه ما، این مورد صدق نمی‌کند و علت نیز همان کمی حجم نمونه در این رده سنی است. در مطالعات دیگر ارتباطی میان تعداد پلاکت‌ها، لنفوسیت‌ها و گلبول‌های سفید با پیش آگهی و زمان بقا وجود نداشته است (۲۲-۲۳) و در مطالعه ما نیز وجود چنین ارتباطی تأیید نشده است؛ اما، این ارتباط بین هموگلوبین و سلول‌های نابالغ با زمان بقا وجود دارد، به طوری که با استفاده از فرمول پیش‌بینی کننده زیر می‌توان زمان بقای بیماران را بر حسب روز محاسبه کرد.

$$\hat{Y} = 81.0353 + 22.5674 * HGB - 6.3590 * LUC + 24.4034 * AGE1$$

مسلم است که محدودیت عمده این مطالعه، بررسی این پارامترها در بیماران فوت شده است. از طرفی سایر عوامل پیش آگهی، مانند عوامل سیتوژنتیک، در این مطالعه منظور نشده‌اند؛ اما با توجه به این که

## منابع

subgroups of adult ALL: analysis of 1200 patients in GMALL study. Blood 2001;98:802a-3a.

4. Hunault M, Harousseau JL, Delain M, Truchan-Graczyk M, Cahn JY, Witz F, et al. Better outcome of adult acute lymphoblastic leukemia after early genoidentical allogeneic bone marrow transplantation (BMT) than after late high-dose therapy and autologous BMT: a GOELAMS trial. Blood 2004;104(10):3028-37.

5. Kantarjian H, Thomas D, O'Brien S, Cortes J, Giles F, Jeha S, et al. Long-term follow-up results of hyperfractionated cyclophosphamide, vincristine, doxorubicin, and dexamethasone (Hyper-CVAD), a

۱. صمدانی فرد حسین، ستوده نیا عبدالحسین. بیماریهای هماتولوژی و انکولوژی (ترجمه). از هریسون، تنسلی، راندولف. اصول طب داخلی هاریسون. چاپ اول. تهران: موسسه انتشاراتی نسل فردا؛ ۱۳۸۴.

2. Annino L, Vegna ML, Camera A, Specchia G, Visani G, Fioritoni G, et al. Treatment of adult acute lymphoblastic leukemia (ALL): long-term follow-up of the GIMEMA ALL 0288 randomized study. Blood 2002;99(3):863-71.

3. Goekbuget N, Arnold R, Buechner T, Ganser A, Ludwig WD, Maschmeyer G, et al. Intensification of induction and consolidation improves only

- distributions and the Bayesian restoration of image. *IEEE Trans Pattern Anal Mach Intell* 1984;6(6):721-41.
17. Mengersen KL, Robert CP, Guihennelle-Jouyoux C. MCMC convergence diagnostics; a review. In: Berger JO, Bernardo JM, David AP, Lindley DV, Smith AFM, editors. *Bayesian Statistics*. New York: Oxford University Press; 1999. p. 415-40.
18. Moussaoui S, Carteret C, Djafari AM. Bayesian Analysis of spectral mixture data using Markov Chains Monte Carlo Methods. *Chemometrics and Intelligent Laboratory Systems* 2006;81(2):137-48.
19. Pui CH, Evans WE. Treatment of acute lymphoblastic leukemia. *N Engl J Med* 2006;354(2):166-78.
20. Takeuchi J, Kyo T, Naito K, Sao H, Takahashi M, Miyawaki S, et al. Induction therapy by frequent administration of doxorubicin with four other drugs, followed by intensive consolidation and maintenance therapy for adult acute lymphoblastic leukemia: the JALSG-ALL93 study. *Leukemia* 2002;16(7):1259-66.
21. Brincker H. Estimate of overall treatment results in acute nonlymphoblastic leukemia based on age-specific rates of incidence and of complete remission. *Cancer Treat Rep* 1985;69(1):5-11.
22. Mertelsmann R, Tzvi Thaler H, To L, Gee TS, McKenzie S, Schauer P, et al. Morphological classification, response to therapy, and survival in 263 adult patients with acute nonlymphoblastic leukemia. *Blood* 1980;56(5):773-81.
23. Hug V, Keating M, McCredie K, Hester J, Bodey GP, Freireich EJ. Clinical course and response to treatment of patients with acute leukemia. *Cancer* 1983;52(5):773-9.
- dose-intensive regimen, in adult acute lymphocytic leukemia. *Cancer* 2004;101(12):2788-801.
6. Larson RA. The U.S. trials in adult acute lymphoblastic leukemia. *Ann Hematol* 2004;83(Suppl 1):S127-8.
7. Linker C, Damon L, Ries C, Navarro W. Intensified and shortened cyclical chemotherapy for adult acute lymphoblastic leukemia. *J Clin Oncol* 2002;20(10):2464-71.
8. Thomas X, Boiron JM, Huguet F, Dombret H, Bradstock K, Vey N, et al. Outcome of treatment in adults with acute lymphoblastic leukemia: analysis of the LALA-94 trial. *J Clin Oncol*. 2004;22(20):4075-86.
9. Qian SS, Stow CA, Borsuk ME. On Monte Carlo Methods for Bayesian inference. *Ecol Modell* 2003;159:269-77.
۱۰. بهرامی محمد. استنباط آماری بیزی در مدل‌های آمیخته (پایان نامه دوره دکتری آمار). گروه آمار دانشگاه شهید بهشتی. ۱۳۷۶.
11. Karli D, Xekalaki E. Choosing initial values for the EM algorithm for finite mixtures. *Comput Stat Data Anal* 2003;41(3-4):577-90.
12. Dempster AP, Laird NM, Rubin D. Maximum Likelihood from incompleteData via the EM algorithm. *J Roy Statist Soc Ser B* 1977;39(1):1-38.
13. Klien JP, Moeschberger ML. *Survival Analysis*. New York: Springer; 1997.
14. Bayarri MJ, Castellanos ME, Morales J. MCMC methods to approximate conditional predictive distributions. *Comput Stat Data Anal* 2006;51(2):621-40.
15. Gelfand AE, Smith AFM. Sampling Based approaches to calculating marginal densities. *J Am Stat Assoc* 1990;85(410):398-409.
16. Geman S, Geman D. Stochastic relaxation, Gibbs



Received: 17.1.2007

Accepted: 2.3.2007

**Effective Factors on Survival Time of the leukemic Patients and Estimating the Mean of Survival Time by Expectation and Maximization Algorithm and Monte Carlo Markov Chains Simulation Method**

Bahrami M PhD\*, Moshkani MR PhD\*\*, Alam Samimi M MD\*\*\*

\* Faculty Member of Department of Statistics, University of Isfahan, Isfahan

\*\* Professor of Statistics, Department of Statistics, Shahid Beheshti University, Tehran

\*\*\* MD, Hematology Oncologist, Isfahan University of Medical Sciences, Isfahan

**Background:****Abstract**

Leukemia is a kind of malignancy blood system which leads to death of human beings in a very short period of time. In this paper, the effective factors on survival time of the acute lymphoblastic leukemia (ALL) patients have been considered to achieve a linear regression model show the relation between the life-time after diagnosis and some explanatory factors.

**Methods:**

In this study, the data of 52 patients died from ALL was used. The designed model contained three variables, hemoglobin, large undifferentiated cell (LUC) and age. According to the data suggesting, a kind of mixture distribution, we considered a mixture model for survival time. Applying the EM-algorithm, we have found the maximum likelihood estimate of mean survival time and the Bayesian estimate of the mean survival time by Monte Carlo Markov Chain method.

**Findings:**

Based on the obtained estimating survival function, we can predict the survival time of the patients and decide about their treatment protocol.

**Conclusion:**

It is suggested that by conducting larger studies and statistical analysis used in this paper, a correlative can be found between clinical & paraclinical findings and the survival time. This model can be used in often kinds of diseases for determining the prognosis.

**Key words:**

**Maximum likelihood estimation, bayesian estimation, bimodal, leukemia, mixture models, survival mean.**

**Page count:**

9

**Tables:**

5

**Figures:**

2

**References:**

23

**Address of Correspondence:**

Mohammad Bahrami PhD, Department of Statistics, University of Isfahan, Isfahan, Iran.

E-mail: bahrami@sci.ui.ac.ir