# کاربرد یک مدل تصادفی برای توزیع دوره پنهانی ویروس ایدز بر پایه شمارش T4 یا cd4

احد ملک زاده[1]


## چکیده

یک مدل ریاضی برای توصیف سطح $T4$، بعنوان یک تابع از زمان عفونت اولیه، در خون شخصی که به $HIV$ مبتلا می‌باشد طراحی شده‌است. برای هر شخص مبتلا به عفونت دو متغیر تصادفی تعریف می‌شود، زمان $T$ لحظه عفونت تا لحظه‌ای که عفونت برای اولین بار ظاهر می‌شود، و متغییر $R$، سطح سلولهای $T4$ در خون در زمان کشف عفونت، به عنوان دومین متغیر تصادفی می باشند. در اینجا $T$ طول دوره پنهانی است. بر اساس مدل تصادفی بکار رفته یک روند گاوسی ایستا با چگالی توأم برای این دو متغیر تصادفی بدست می‌آوریم و سپس توزیع شرطی $T$ به شرط $R$ را بعنوان توزیع مورد نظر بدست خواهیم آورد. نشان خواهیم داد که اگر $T$ دارای توزیع حاشیه‌ای نمائی باشد آنگاه توزیع شرطی $T$ یک توزیع نرمال بریده خواهد بود. این تئوری برای داده‌های جمع‌آوری شده از بیماران شهر نیویورک بکار گرفته شده‌است. با استفاده از داده‌های واقعی جمع‌آوری شده مربوط به بیماران شهر کرمانشاه، تئوری مذکور را بکارگرفته و این مدل را روی آن برازنده می‌نمائیم.


**واژه‌های کلیدی:** توزیع نرمال بریده شده، توزیع نمائی، چندجمله‌ای هرمیت، توزیع پیشین، توزیع پسین، فرآیند گاوسی

## ۱. مقدمه

فرض کنید جواب آزمایش شخصی که قبلاً برای عفونت ویروس $HIV$ مثبت نشده بود، حالا این آزمایشها را انجام داده و مشخص شده است که دارای عفونت می‌باشد. در این زمان سطح $T4$ درون خون این شخص اندازه‌گیری می شود. حال درباره‌ی مدت زمان سپری شده از زمان عفونت اولیه بر اساس شمارش $T4$ها چه استنباطی می توانیم داشته باشیم؟ در همه موارد بجز موارد استثناء زمان عفونت در بیماران ناشناخته است. بنابراین سطح $T4$ و زمان سپری شده از عفونت اولیه را بعنوان متغیرهای تصادفی در نظر گرفته و با فرض‌های فراهم شده توزیع توأم این دو متغیر تصادفی را برآورد کنیم سپس توزیع زمان عفونت را با توجه به سطح $T4$ای که در زمان کشف بیماری بدست آمده شرطی خواهیم کرد.

تعداد $T4$ از زمان اولیه عفونت دارای یک روند نزولی می‌باشد. تجزیه و تحلیل ما بر اساس یک مقایسه از کمترین سطح $T4$ با حالت نرمال در غیاب ویروس $HIV$ پایه‌ریزی شده است. فرض کنید $X(T)$ ($T$ یک متغیر زمان است) نمایشگر تعداد $T4$ در شخص $HIV$ داری باشد که به‌صورت تصادفی انتخاب شده است. $X(t)$ را بعنوان تابع نمونه‌ای از یک روند ایستای[2] تصادفی و روند سپری شدن از زمان عفونت ویروس $HIV$ را بعنوان یک عامل نمائی در نظر می‌گیریم یعنی اگر عفونت ویروس $HIV$ در زمان $t_0$ اتفاق بیفتد، سطح $T4$ در زمان $t > t_0$ از $X(t)$ در حالت طبیعی به $X(t)\exp\{-\partial(t-t_0)\}$، که در آن $\partial > 0$ نرخ کاهش تعداد $T4$ در واحد زمان است، تغییر می‌کند.

به‌جای تلاش جهت مشخص کردن زمان $t_0$ برای هر شخص، توزیع زمان سپری‌شده از عفونت را برای جامعه‌ای از اشخاصی که برای اولین بارآزمایش شده‌اند و مشخص شده که دارای عفونت می‌باشند برآورد می‌کنیم. برای چنین افرادی ما عدد $T$ را بعنوان زمان اولین آزمایش مثبت منهای زمان عفونت تعریف می‌کنیم. هرچند که قابل مشاهده نیست اما $T$ خوش تعریف است. $T$ را بعنوان یک متغیر تصادفی نامنفی دارای تابع چگالی نامشخص $q(t)$ در نظر می‌گیریم و فرض می‌کنیم که $T$ مستقل از روند تصادفی $X(t)$ می باشد. آنالیز با این فرض ساده شده و مدل حاصل بعنوان یک استاندارد برای باقی روند و مدل‌های آینده قابل بکارگیری است. بنا به تعریف بالا سطح $T4$ در $T$ واحد زمانی پس از $t_0$ (زمان عفونت) با $X(t_0+T)\exp(-\partial T)$ نمایش داده می‌شود. برای $K > 1$ اگر $t_1, t_2, t_3, \ldots, t_{K-1}$ مدت زمان

۱- دانشجوی کارشناسی ارشد بخش آمار دانشگاه شیراز

۲- منظور از اولین ملاقات آزمایشی است، که عفونت برای اولین بار کشف می‌شود و منظور از Kامین ملاقات K-1 امین آزمایش پس از زمان کشف بیماری ایدز در شخص مبتلا می‌باشد.

بین ۱ − $K$ ملاقات[۲] (آزمایش) بعدی در کلینیک یا بیمارستان را نمایش دهد.

آنگاه سطوح $T4$ مشاهده شده در $K$ ملاقات بصورت

$$X(t_0+T)\exp\{-\partial T\},$$
$$X(t_0+t_1+T)\exp\{-\partial(T+t_1)\},\ldots,$$
$$X(t_0+T+t_1+\ldots+t_{K-1})\exp\{-\partial(T+t_1+\ldots+t_{K-1})\} \quad (۱.۱)$$

قابل نمایش می‌باشد. توزیع بردار بالا مستقل از $t_0$ است، بنابراین در آنالیز آماری کافی است فرض کنیم $t_0 = 0$ یعنی زمان عفونت اولیه برای هر شخص بعنوان زمان مبدأ در نظر گرفته می‌شود. (۱.۱) بصورت زیر جایگزین می‌شود.

$$X(T)\exp\{-\partial T\},\ X(t_1+T)\exp\{-\partial(T+t_1)\},\ldots,$$
$$X(T+t_1+\ldots+t_{K-1})\exp\{-\partial(T+t_1+\ldots+t_{K-1})\}$$

در عمل تنها به موارد $k = 2$ توجه می‌کنیم. بنابراین داده‌های ما شامل دو جفت $X(t_1+T)\exp\{-\partial(T+t_1)\}$ و $X(T)exp\{-\partial(T)\}$ است که از بیماران انتخابی بدست می‌آید. توجه کنید متغیر تصادفی $X(0)$ معرف سطح $T4$ در لحظه عفونت می‌باشد.

$q(t)$ را توزیع پیشین و توزیع شرطی $T$ را توزیع پسین گوییم. برآورد توزیع پسین موضوع این مقاله می‌باشد. برای این کار ابتدا توزیع پیشین را از داده‌های مشاهده شده

$$X(t_1+T)exp\{-\partial(T+t_1)\}, X(T)\exp\{-\partial(T)\}$$

برآورد و سپس با استفاده از قانون بیز توزیع شرطی یا توزیع پسین را بدست می‌آوریم.

این مقاله به هشت بخش تقسیم بندی می‌گردد که مقدمه در بخش اول آن است. در بخش دوم فرضیات مدل و در بخش ۳ نشان خواهیم داد چگونه گشتاورهای توزیع پیشین را می‌توان از گشتاورهای هرمیت مشاهده شده از سطوح $\log(T4)$ در اولین ملاقات برآورد کرد. یک فرمول ساده برای برآورد همبستگی بین $T4$ و $X(T)$ را نیز پیشنهاد می‌کنیم. در بخش ۴ توزیع پسین را با فرض توزیع پیشین نمائی بدست می‌آوریم و نشان می‌دهیم، توزیع پسین یک توزیع نرمال بریده شده‌است. در بخش ۵ برآورد پارامترهای مدل آمده‌است. بخش ۶ مربوط به داده‌های جمع‌آوری شده شهر نیویورک و بخش ۷ نیز این تئوری روی داده‌های شهر کرمانشاه پیاده می‌شود. نتایج حاصل از مقایسه توزیع پسین بدست‌آمده از بیماران این دو شهر در بخش ۸ بیان خواهدشد.

## ۲. فرضیات ریاضی ویژه

اگر فرض کنیم $X(T)$ یک روند ایستای تصادفی دارد آن‌را می‌توان بصورت زیر نشان داد.

$$X(t) = \exp\{z(t)\} \quad (۱.۲)$$

که $Z(t)$ ایستا و دارای توزیع نرمال می‌باشد (با توجه به فرض لگ نرمال برای مدلهای کلی در متغیرهای بیولوژیکی کرامر(۱۹۴۶ صفحه ۲۰۰)). با توجه به فرض ایستائی، $E\{Z(t)\}$ و $E\{Z^2(t)\}$ ثابت هستند. تعریف می‌کنیم:

$$r(t,s) = E\{Z(s)Z(t)\} - E\{Z(t)\}E\{Z(s)\}$$

که $r(t,s)$ تابعی از $|s-t|$ است و قرار می دهیم

$$\sigma^2 = \mathrm{var}\{Z(t)\} \text{ و } \mu = E\{Z(t)\} \quad (۲.۲)$$

با توجه به این مقدمه و رابطه (۱,۲) داریم $X(t)\exp\{-\partial t\} = \exp\{z(t)-\partial t\}$، با قرار دادن $R(t) = Z(t) - \partial t$، که $R(t)$ دارای روند گاوسی با واریانسی همانند $Z(t)$ و میانگین $\mu - \partial t$ است. تئوری توزیعی مورد بحث در سه متغیر

۱. $R(0) = Z(0) = log(T4)$ در لحظه عفونت
۲. $R(T) = log(T4)$ در اولین ملاقات
۳. $R(t_1+T) = log(T4)$ در دومین ملاقات

متمرکز شده است. مدل دارای سه پارامتر نامعلوم $\mu, \sigma^2, \partial$ وتابع کوواریانس نامشخص $r(t)$ و یک توزیع پیشین برای $T$ می‌باشد.

$T$ مستقل از روند $Z(T)$ است زیرا مستقل از روند $X(T)$ فرض شده بود. بنابراین توزیع توأم متغیرهای $R(T)$ و $R(t_1+T)$، بشرط $T = t$ با توزیع توأم غیرشرطی $R(T)$ و $R(t_1+T)$ برابر می‌باشد و دارای توزیع نرمال دو متغیره با بردار میانگین

$$\begin{pmatrix} \mu - \partial t \\ \mu - \partial t - \partial t_1 \end{pmatrix} \quad (۳.۲)$$

و ماتریس واریانس-کوواریانس

$$\begin{bmatrix} \sigma^2 & r(t) \\ r(t) & \sigma^2 \end{bmatrix} \quad (۴.۲)$$

می‌باشد.

$\overline{R}(t_1) = \{R(T) - R(t_1+T)\}/t_1$ میانگین تغییرات $\log(T4)$ بین اولین و دومین ملاقات بیماران در واحد زمانی را نشان

می‌دهد. $\overline{R}(t_1)$ به شرط معلوم بودن مقدار $T$ دارای توزیع نرمال است یعنی

$$\overline{R}(t_1) \sim N\left(\partial,\ 2\{\sigma^2 r(t_1)\}/t_1^2\right) \qquad (۵.۲)$$

با توجه به فرضیات در مورد $Z(t)$، که $R(0)$ دارای توزیع نرمال با میانگین $\mu$ و واریانس $\sigma^2$ می‌باشد و توزیع شرطی بالا، نتیجه می‌گیریم که $R(T)$ دارای توزیع غیر شرطی شبیه متغیر $Z(0)-\partial T$ است، بطوری‌که $Z(0)$ و $T$ مستقل از یکدیگرند. از این نمایش یک فرمول صعودی برای ضریب همبستگی بین $R(T)$ و $T$ بدست می‌آید.

$$corr(R(T),T) = -\left\{\frac{var(\partial T/\sigma)}{1+var(\partial T/\sigma)}\right\}^{1/2} \qquad (۶.۲)$$

در بخش ۵ نشان می‌دهیم که پارامترهای $\mu$، $\sigma^2$ و $\partial$ بسادگی قابل برآوردند. در ادامه به‌منظور برآورد چگالی‌های پیشین و پسین از $T$ فرض کنید که $\mu$ ، $\sigma^2$ و $\partial$ معلوم می‌باشند. با انجام تبدیلات مناسب بر روی متغیرهای تصادفی $R(T)$ و $T$ بصورت $\{R(T)-\mu\}/\sigma$ و $\partial T/\sigma$ (۲،۲) را می‌توان نوشت برابر با

$$\{R(T)-\mu\}/\sigma = \{Z(t)-\mu\}/\sigma - \partial t/\sigma$$

که $\{Z(T)-\mu\}/\sigma$ دارای میانگین صفر و واریانس ۱ می‌باشد. اگر فرض کنیم که $q(t)$ چگالی متغیر تبدیل یافته $\partial T/\sigma$ باشد آنگاه بنا به(۳،۲) و (۴،۲) توزیع شرطی $\{R(T)-\mu\}/\sigma$ با مقدار داده شده $\partial T/\sigma = t$، توزیع نرمال با میانگین $-t$ و واریانس ۱ می‌باشد. چگالی توأم $\partial T/\sigma$ و $\{R(T)-\mu\}/\sigma$ را می‌توان نوشت.

$$\phi(x+t)q(t) \qquad (۷.۲)$$

که چگالی حاشیه‌ای $\{R(T)-\mu\}/\sigma$ برابراست با

$$\int_0^\infty \phi(x+t)q(t)dt \qquad (۸.۲)$$

و چگالی پسین $\partial T/\sigma$ یعنی چگالی شرطی $\partial T/\sigma$ بشرط $\{R(T)-\mu\}/\sigma = x$ بصورت زیر می‌باشد.

$$\frac{\phi(x+t)q(t)}{\left\{\int_0^\infty \phi(x+s)q(s)dS\right\}} \qquad (۹.۲)$$

## ۳. شناسائی چگالی پیشین

فرض کنید

$$g(x) = \int_0^\infty \phi(x+t)q(t)dt \qquad (۱.۳)$$

یک تابع چگالی است و بطور یکتا با $q(t)$ مشخص می‌شود، و برعکس آن نیز درست است یعنی اگر $g(x)$ یک چگالی قابل نمایش بفرم (۱.۳) برای تعدادی چگالی $q(t)$ باشد آنگاه چگالی $q(t)$ یکتاست. برای اثبات اگر $\hat{g},\hat{\phi}$ و $\hat{q}$ را به ترتیب تبدیل فوریه $g,\phi$ و $q$ در نظر بگیریم آنگاه (۱. ۳) قابل نمایش به صورت $\hat{g}(y) = \hat{q}(-y)\times\hat{\phi}(y)$ می‌باشد در حالیکه $\hat{\phi}(y) = \exp(-\frac{1}{2}y^2)$، این نشان دهنده رابطه‌ی یکتا بین $q(t)$ و $g(x)$ است.

با استفاده از تعریف چند جمله‌ای هرمیت رابطه بین گشتاورهای $q(t)$ و $g(x)$ را نشان می‌دهیم[۱].

طبق تعریف $H_0(x) \equiv 1$ و برای $m \geq 1$ $H_m(x)$ چند جمله‌ای از درجه $m$ می‌باشند که در رابطه زیر صدق می‌کند.

$$\left(\frac{d}{dx}\right)^m \phi(x) = (-1)^m H_m(x)\phi(x) \qquad (۲.۳)$$

برای نمونه چهار گشتاور اول $H_m(x)$ را ذکر می‌کنیم.

$$\frac{d\phi(x)}{dx} = (-1)x\phi(x) \Rightarrow H_1(x) = x$$

$$\frac{d^2\phi(x)}{dx^2} = (-1)^2(x^2-1)\phi(x) \Rightarrow H_2(x) = x^2-1$$

$$\frac{d^3\phi(x)}{dx^3} = (-1)^3(x^3-3x)\phi(x)$$

$$\Rightarrow H_3(x) = x^3-3x$$

$$\frac{d^4\phi(x)}{dx^4} = (-1)^4(x^4-6x^2-1)\phi(x)$$

$$\Rightarrow H_4(x) = x^4-6x-1$$

**قضیه ۳. ۱:** اگر $H_m(x)$ تابعی باشد که در شرط (۳. ۲) صدق کنند آنگاه خواهیم داشت

$$\int_{-\infty}^{\infty} H_m(z)\phi(z-t)dz = t^m \qquad (۳.۳)$$

برای اثبات از روش جزء به جزء می توان استفاده نمود

اگر در رابطه (۳. ۳) $t$ را با Y جایگذاری کنیم لم زیر براحتی اثبات می شود.

**لم ۳. ۱:** اگر X و Y دو متغیر مستقل باشند بطوریکه X دارای یک چگالی نرمال باشد و $E(|Y|^m) < \infty$ $(m \geq 1)$ آنگاه $E\{H_m(X+Y)\} = E(Y^m)$.

با توجه به بخش ۲ مسی نویسیم $(R-\mu)/\sigma = (Z-\mu)/\sigma + \partial T/\sigma$ که R متغیری تصادفی

از $Log(T4)$ و $\{Z(T)-\mu\}/\sigma$ دارای توزیع نرمال استاندارد است که $T$ نیز مستقل از $Z$ است. بنا به لم ۳-۱با قرار دادن $X=(Z-\mu)/\sigma$ و $Y=-\partial T/\sigma$ داریم که

$$E\left\{H_m\left(\frac{R-\mu}{\sigma}\right)\right\} = E\left\{\left(-\frac{\partial T}{\sigma}\right)^m\right\} \quad (m=0,1,2,\ldots) \qquad (۴.۳)$$

حال اگر $R_1,R_2,\ldots,R_n$ نشان دهنده مقادیر نمونه‌ای از $R$ باشند آنگاه $\frac{1}{n}\sum_{g=1}^{n}H_m(\frac{R_g-\mu}{\sigma})$ یک برآورد از $(-1)^m E\left\{\left(\frac{\partial T}{\sigma}\right)^m\right\}$ می‌باشد.

## ٤. چگالی پسین با توجه به چگالی پیشین نمائی

فرض کنید چگالی پیشین در(۱,۳) یک چگالی نمائی با $m>0$ باشد، یعنی

$$q(t)=\begin{cases} m^{-1}\exp\{-t/m\} & t>0 \\ 0 & t\le 0\end{cases} \qquad (۱.۴)$$

پس تابع چگالی توأم (۷,۲) برای $t>0$ برابر است با $\phi(x+t)m^{-1}\exp\{-t/m\}$ که می توان نوشت

$$f(x,t)=\sqrt{2\pi}\phi(x)\phi(t)m^{-1}\exp\{-t(m^{-1}+x)\} \qquad (۲.۴)$$

**توجه۴. ۱:** اگر $r$ مقدار مشاهده شده از لگاریتم تعداد cd4 در اولین ملاقات باشد، آنگاه $x,r$ در رابطه $r=\mu+\sigma x$ صدق می‌کنند.

**توجه۴. ۲:** اگر مدت زمان دوره پنهانی ویروس ایدز $t^*$ مشاهده شود آنگاه $t=\frac{\delta t^*}{\sigma}$.

چگالی حاشیه‌ای(۸,۲) برابر است با

$$f(x)=\sqrt{2\pi}m^{-1}\phi(x)\int_0^\infty \phi(t)\exp\{-t(m^{-1}+x)\}dt \qquad (۳.۴)$$

چگالی پسین(۹,۲) را نیز بصورت زیر بدست می‌آوریم.

$$f(t/x)=\frac{\phi(t)\exp\{-t(m^{-1}+x)\}}{\int\phi(s)\exp\{-s(m^{-1}+x)\}ds} \quad (t>0) \qquad (۴.۴)$$

حدود انتگرال بین صفر تا بینهایت می‌باشد. اگر در انتگرال نشان داده شده در(۴.۳) قرار دهیم

$$\exp\{-sr\}\phi(s)=\exp\{\tfrac{1}{2}r^2\}\phi(s+r)$$

خواهیم داشت:

$$\int_0^\infty \exp\{-sr\}\phi(s)ds=\exp\{\tfrac{1}{2}r^2\}\int_r^\infty \phi(s)ds \qquad (۵.۴)$$

با قرار دادن $\psi(z)=\int_z^\infty \phi(s)ds$، چگالی حاشیه‌ای (۳,۴) بدست می آید.

$$f(x)=\sqrt{2\pi}m^{-1}\phi(x)\exp\{\tfrac{1}{2}(m^{-1}+x)^2\}\psi(m^{-1}+x) \qquad (۶.۴)$$

بنابراین چگالی شرطی(۴,۴) مساوی تابع زیر است

$$f(t/x)=\frac{\phi(m^{-1}+t+x)}{\psi(m^{-1}+x)} \quad (t>0) \qquad (۴,۷)$$

که

$$f(x)=m^{-1}\exp\{\tfrac{1}{2}m^{-2}\}\exp\{x/m\}\psi(m^{-1}+x) \qquad (۸.۴)$$

توجه کنید که (۷.۴) چگالی نرمال بریده‌شده یک متغیر تصادفی است.

میانگین و واریانس چگالی پسین(۴,۷) را می توان بصورت زیر نوشت:

$$M(x)=\frac{\phi(m^{-1}+x)}{\psi(m^{-1}+x)}-(m^{-1}+x)$$

$$V(x)=\frac{4+(m^{-1}+x)^2}{4}-\left\{\frac{\phi(m^{-1}+x)}{\psi(m^{-1}+x)}-\frac{(m^{-1}+x)}{2}\right\}^2 \qquad (۹.۴)$$

می‌توان نشان داد:

$$dM/dx=-V(x) \qquad (۱۰.۴)$$

یک نتیجه خاص از (۱۰,۴) این است که تابع $M(x)$ یک تابع غیرصعودی از $x$ است.

## ۵. برآورد $\mu$ ، $\sigma^2$ ، $\partial$

بنا به رابطه (۲,۲) پارامترهای $\mu,\sigma^2$ بترتیب واریانس و میانگینی از $\log(T4)$ در لحظه عفونت می‌باشند. پارامترها را می‌توان با استفاده

از نمونه‌ای از افراد این جامعه برآورد کرد. اگر $Z$ دارای یک توزیع نرمال $(\mu,\sigma^2)$ باشد، آنگاه

$$E(e^z)=\exp\{\mu+\frac{1}{2}\sigma^2\},\ E(\exp\{2z\})=\exp\{2\mu+2\sigma^2\}$$

$\mu$ و $\sigma^2$ در معادله‌های زیر نیز صدق می کنند.

$$\mu=2\log E(\exp\{z\})-\frac{1}{2}\log E(\exp\{2z\})$$

$$\sigma^2=\log E(\exp\{2z\})-2\log E(\exp\{z\}).$$

در رابطه (۵.۲) پارامتر $\partial$ را اختلاف مورد انتظار از لگاریتم اولین و دومین شمارش $\log(T4)$ تقسیم بر طول بازه زمانی $t_1$ تعریف می‌کنیم یعنی

$$\{R(T)-R(T+t_1)\}/t_1 \qquad (۱.۵)$$

واریانس متغیر تصادفی (۱.۵) از فرمول دوم(۵.۲) که شامل پارامترهای مجهول $\sigma^2$ ، $r(t_1)$ و اختلاف زمانی $t_1$ می‌باشد بدست می‌آید. واریانس الزاماً برای همه مقادیر(۱. ۵) یکسان نیست زیرا برای $t_1$های مختلف، ممکن است $r(t_1)$ متفاوت باشد. اگر پارامترها مشخص باشند یک متوسط وزنی متناسب از مشاهدات (۵،۱)، برآوردگری نااریب با کمترین واریانس را مهیا می‌کند. بدلیل اینکه پارامترها معلوم نیستند، برآورد نااریب نمونه‌ای از $\partial$ را متوسط مشاهدات(۵،۱) قرار می‌دهیم.

## ۶. کاربرد داده‌های جمع‌آوری شده از بیماران تزریقی در شهر نیویورک

روشهای بخش ۳ و۵ برای برآورد پارامترهای مدل تعریف شده در بخش ۱ و۲، که برای داده‌های جمع‌آوری شده از مصرف‌کنندگان داروهای تزریقی در ترک اعتیاد و برنامه‌های داروئی ترک اعتیاد در شهر نیویورک که بوسیله لارلیس[1]، یانگ[2]، دربین[3] و همچنین مارمور[4] تشریح شده است استفاده می‌شود. اولین ملاقات نمونه‌ها در بین سال‌های ۱۹۸۴ تا ۱۹۸۵ ثبت شده و دومین ملاقات نیز برای یکسال بعد برنامه ریزی شده بود. بیشتر افراد برای ملاقات بعدی بازنگشتند و افرادی نیز که برگشتند پس از موعد مقرر یکساله برای دومین ملاقات حضور پیدا کردند. $\mu$ و $\sigma^2$ بر اساس داده‌های بدست آمده از افراد $HIV$دار که ۱۹۱ نفر بودند برآورد شده است. برآورد $\partial$ و گشتاورهای $T$ از داده‌هائی بدست می‌آید که افراد برای ملاقات دوم نیز بازگشته‌اند، تعداد آنها ۵۹ نفر است.

برآورد میانگین و واریانس بر پایه نمونه ۱۹۱ نفری از افراد $HIV$ دار برابر است با

$$\hat{\mu}=۶/۹۶۶ \quad و \quad \hat{\sigma}=۰/۳۵۴ \qquad (۱.۶)$$

یک میانگین نمونه‌ای از نسبت(۵،۱) در طول اولین و دومین ملاقات برابر $\hat{\partial}=۰/۰۳۳۵$ است. در اینجا واحد اندازه‌گیری زمان ماه می‌باشد.

حال چگونگی برآورد گشتاورهای متغیر تصادفی $T$ را بررسی می‌کنیم. در بخش ۳ فرض شد $R$ نمایانگر یک مقدار نمونه‌ای از $\log(T4)$ باشد. n را برابر با اندازه نمونه قرار می‌دهیم. با جایگذاری مقادیر بدست آمده مربوط به میانگین و واریانس در(۱،۶) چهار گشتاور اول نمونه‌ای $(R-\mu)/\sigma$ را بدست می‌آوریم.

$$\frac{1}{n}\sum(\frac{R-\mu}{\sigma})^2=۱۱/۰۱۲۵ \quad \frac{1}{n}\sum(\frac{R-\mu}{\sigma})=-۲/۳۷۴۴$$

$$\frac{1}{n}\sum(\frac{R-\mu}{\sigma})^3=-۸۲/۹۶۲$$

$$\frac{1}{n}\sum(\frac{R-\mu}{\sigma})^4=۹۲۱/۶۶ \qquad (۲.۶)$$

با توجه به (۲،۳) و فرمول های بدست آمده برای $H_j(x)(j=۱,...,۴)$، چهار گشتاور اول $\partial T/\sigma$ را برآورد می‌کنیم

$$-\frac{1}{n}\sum H_1(\frac{R-\mu}{\sigma})=۲/۳۷۴۴$$

$$\frac{1}{n}\sum H_2(\frac{R-\mu}{\sigma})=۱۰/۰۱۲۵$$

$$-\frac{1}{n}\sum H_3(\frac{R-\mu}{\sigma})=۷۵/۸۳۹$$

$$\frac{1}{n}\sum H_4(\frac{R-\mu}{\sigma})=۸۵۷/۹۲۵ \qquad (۳.۶)$$

با استفاده از دو رابطه اول در(۳،۶) واریانس برآورد شده $\partial T/\sigma$ برابر با $۱۰/۰۱۲۵-(۲/۳۷۴۴)^2=۴/۳۷۴۷$ می‌باشد. بنابراین برآورد ما از ضریب همبستگی $R(T)$ و $T$ در(۲،۶) برابر با $-۰/۹۰۲۲$ است. این ضریب همبستگی منفی، بسیار نزدیک به ۱- است. بدین معنی که یک رابطه خطی منفی قوی بین $R(T)$ و $T$ وجود دارد.

حال نشان می‌دهیم این گشتاورها کاملاً نزدیک به گشتاورهای توزیع نمائی با میانگین ۲/۳۷۴۴ هستند که عبارتند از

---

۱- Larlais
۲- Yang
۳- Dirbin
۴- Marmor

$$\alpha_1 = ۲/۳۷۴۴,\ \ \alpha_2 = ۱۱/۲۷۶$$
$$\alpha_3 = ۸۰/۳۱۸,\ \ \alpha_4 = ۷۶۲/۸۳ \qquad (۴.۶)$$

مقایسه(۳،۶) و (۶،۴) الزاما اشاره به این ندارد که $\partial T/\sigma$ دارای توزیع نمائی است اما می‌توان آنرا بعنوان یک تقریب پذیرفت. یکتائی بین $q$ و $g$ اشاره به این دارد که $q$ دارای چگالی(۱،۴)است اگر و تنها اگر چگالی $g$ از(۴،۸) آمده باشد. اگر چگالی نمائی را بعنوان چگالی $\partial T/\sigma$ بپذیریم آنگاه می‌توان مدل (۴،۸) را بعنوان چگالی $(R-\mu)/\sigma$ پذیرفت. حال با قرار دادن $m = ۲/۳۷۴۴$ چگالی حاشیه‌ای زیر برای $(R-\mu)/\sigma$ نتیجه می شود.

$$۰/۴۶۰\exp\{x/۲.۳۷۴۴\}\psi(x+۰/۴۲۱) \qquad (۵.۶)$$

نمودار ۱ هیستوگرام مقادیر $(R-\mu)/\sigma$ را برای مواردی که برای اولین بار جواب آزمایش آنها مثبت بوده، نشان می‌دهد، بدون توجه به اینکه برای ملاقات بعد، برگشته‌اند یا خیر و با چگالی(۶ .۵) مقایسه شده‌است.

مطابق فرمول(۴ .۷) و میانگین $m^{-1} = ۰/۴۲۱$ چگالی پسین $\partial T/\sigma$ با توجه به مقدار داده شده $R = \mu + \sigma x$ برابر است با:

$$\frac{\phi(t+x+۰/۴۲۱)}{\psi(x+۰/۴۲۱)} \qquad (۶.۶)$$

میانگین و واریانس این چگالی نیز بوسیله فرمول (۹،۴) و با جایگذاری $m = ۲/۳۷۴۴$ بدست می‌آید.

## ۷. تحلیل داده‌های مرتبط با بیماران مبتلا به ویروس ایدز در کرمانشاه

داده های مورد نیاز و مرتبط با روند کار را از شبکه بهداشت و درمان شهرستان کرمانشاه تهیه کردیم. داده های جمع آوری شده متعلق به ۴۷۷ نفر بود که برای اولین بار(در زمان تشخیص بیماری) میزان cd4 آنها اندازه گیری شده است و از این تعداد ۲۱۶ نفر آنها برای دومین بار نیز این مشخصه از خون آنها اندازه گیری شده بود. مابقی بدلائل مختلف آزمایش تشخیص مقدارcd4 آنها برای دومین بار انجام نشده است. این آزمایشات در بین سالهای ۷۹، ۸۰، ۸۱ و ۱۳۸۲ از بیماران به عمل آمده است.

اطلاعات بدست آمده در چهار ستون در نرم افزارSPSS ذخیره شده است. ستون اول و دوم مربوط به تاریخ اولین و دومین روز آزمایش افراد بیمار می باشد، که با نام first وsecond ذخیره شده است. ستون سوم و چهارم مقادیرcd4 های بدست آمده از این آزمایشها است. که با نامهای cd41 و cd42 ذخیره شده است.

با توجه به مقادیر بدست آمده از بیماران کرمانشاه $\mu$ ، $\sigma^2$ را برآورد می کنیم

$$\hat{\sigma} = \sqrt{\ln(۲۴۳۰۲۱/۴۰۶۷) - ۲\times\ln(۴۰۸/۴۳۸۲)} = ۰/۲۳۷۱۴۷ \qquad (۱.۷)$$
$$\hat{\mu} = ۲\times\ln(۴۰۸/۴۳۸۲) - ۰/۵\times\ln(۲۴۳۰۲۱/۴۰۶۷) = ۵/۸۲۴$$

برای بدست آوردن توزیع پیشین نیاز به برآورد گشتاورهای مربوط به توزیع پسین داریم. که با توجه به داده های بدست آمده داریم.

$$\sum_{i=1}^{۲۱۶}\frac{\left(\frac{(\log(cd۴۲)-۵/۸۲۴)}{۰/۲۳۷۱۴۷}\right)}{۲۱۶} = -۲/۳۷۶۸$$
$$\sum_{i=1}^{۲۱۶}\frac{\left(\frac{(\log(cd۴۲)-۵/۸۲۴)}{۰/۲۳۷۱۴۷}\right)^2}{۲۱۶} = ۱۱/۸۲۶۳$$
$$\sum_{i=1}^{۲۱۶}\frac{\left(\frac{(\log(cd۴۲)-۵/۸۲۴)}{۰/۲۳۷۱۴۷}\right)^3}{۲۱۶} = -۶۱/۳۸۵۱$$
$$\sum_{i=1}^{۲۱۶}\frac{\left(\frac{(\log(cd۴۲)-۵/۸۲۴)}{۰/۲۳۷۱۴۷}\right)^4}{۲۱۶} = ۷۵۰/۱۹۷۰۴ \qquad (۲.۷)$$

مقادیر چهارگشاور اول را با استفاده از فرمول های روش هرمیت نیز برآورد می کنیم.

$$-a = ۲/۳۷۶۸ \quad a^2 - ۱ = ۱۰/۸۲۶۳$$
$$-(a^3 - ۳\times a) = ۶۰/۲۵۸۳$$
$$a^4 - ۶a + ۱ = ۷۵۶/۱۲۵۰$$

که در فرمول های بالا

$$a^i = \frac{۱}{n}\sum_{i=1}^{۲۱۶}\left(\frac{\log(cd۴۲)-۵/۸۲۴}{۰/۲۳۷۱۴۷}\right)^i$$

می باشد.

گشتاورهای یک تابع چگالی نمائی با میانگین ۲/۳۷۶۸ بصورت زیر است.

$$\alpha_1 = ۲/۳۷۶۸,\ \alpha_2 = ۱۱/۲۹۸۳$$
$$\alpha_3 = ۸۰/۵۶۱۸,\ \alpha_4 = ۷۶۵/۹۱۷۱ \qquad (۳.۷)$$

در مقایسه مقادیر گشتاورهای تابع چگالی نمائی با میانگین برآورد شده ۲/۳۷۶۸ با مقادیر برآورد شده مربوط به گشتاورهای تابع چگالی پیشین، می توان گفت داده های ما نیز دارای چگالی نمائی هستند.

## ۷. ۱ برآورد تابع چگالی پسین

قبل از برآورد کردن تابع چگالی پسین نیاز است که تابع چگالی پیشین را برآورد کنیم و با توجه به برآوردهای انجام شده برای چهار گشتاور اول و مقادیر واقعی آنها (با توجه به میانگین برآوردی حاصل) در می یابیم که تابع چگالی پیشین ما تقریباً دارای توزیع نمائی با میانگین ۲/۳۴۶۷ است.

در نتیجه تابع چگالی حاشیه ای مربوطه بصورت زیر است:

$$۰/۴۲۰۷\psi \exp\{x/۲/۳۷۶۸\}(x+۰/۴۲۰۷) \qquad (۴.۷)$$

نمودار ۲ هیستوگرام مقادیر $(R-\mu)/\sigma$ را برای افرادی که برای اولین بار جواب آزمایش آنها مثبت بوده را نشان می‌دهد بدون توجه باینکه برای ملاقات بعد، برگشته‌اند یا خیر. در اینجا ۴۷۷ بیمار وجود دارد.

مطابق با فرمول(۷,۴) بخش اول و میانگین $m^{-1} = ۰/۴۲۰۷$ چگالی پسین $\partial T/\sigma$ با توجه به $R = \mu + \sigma x$ عبارتست از

$$\frac{\phi(t+x+۰/۴۲۰۷)}{\psi(x+۰/۴۲۰۷)} \qquad (۵.۷)$$

میانگین و واریانس این چگالی بوسیله فرمول (۱۱,۴) و (۱۲,۴) بخش اول بدست می‌آید یعنی:

$$M(x) = \frac{\phi(۰/۴۲۰۷+x)}{\psi(۰/۴۲۰۷+x)} - (۰/۴۲۰۷+x)$$

$$V(x) = ۱ + \frac{(۴۲/۰۷+x)^۲}{۴} - \left\{\frac{\phi(۰/۴۲۰۷+x)}{\psi(۰/۴۲۰۷+x)} - \frac{(۰/۴۲۰۷+x)}{۲}\right\}^۲ \qquad (۶.۷)$$

مقدار برآورد شده $\delta$ نیز برابر است با $\delta = ۰/۰۳۸۵$. همچنین واریانس برآورد شده $\partial T/\sigma$ برابر است با ۵/۱۸۵۱ = (-۲/۳۷۶۸)$^۲$ - ۱۰/۸۳۴۳. بنابراین برآورد ما از ضریب همبستگی $R(T)$ و $T$ برابر با ۰/۹۱۵۵- می باشد. این نشان دهنده رابطه خطی قوی و معکوس بین $R(T)$ و $T$ می باشد.

## ۸. نتایج

در این بخش اطلاعات و نتایج بدست آمده از دو شهر را با یکدیگر مقایسه می‌کنیم.

برآورد پارامترهای مجهول مدل بوسیله ۱۹۱ بیمار شهر نیویورک بصورت زیر می‌باشد.

$$\hat{\mu}_۱ = ۶/۹۶۶, \hat{\sigma}_۱ = ۰/۳۵۴$$
$$\hat{\alpha}_۱ = ۲/۳۷۴۴, \hat{\delta}_۱ = ۰/۰۳۵۲ \qquad (۱.۸)$$

تابع چگالی پسین برای این داده‌ها بفرم $\frac{\phi(t+x+۰/۴۲۱)}{\psi(x+۰/۴۲۱)}$ است که مقدار امید ریاضی و واریانس مربوط به آن برابر است با:

$$M_۱(x) = \frac{\phi(x+۰/۴۲۱)}{\psi(x+۰/۴۲۱)} - (x+۰/۴۲۱)$$

$$V_۱(x) = ۱ + \frac{(x+۰/۴۲۱)^۲}{۴} - \left\{\frac{\phi(x+۰/۴۲۱)}{\psi(x+۰/۴۲۱)} - \frac{(x+۰/۴۲۱)}{۲}\right\}^۲ \qquad (۲.۸)$$

در بررسی مدل مربوط به ۴۷۷ بیمار شهر کرمانشاه نیز نتایج بصورت زیر می‌باشد.

$$\hat{\mu}_۲ = ۵/۸۲۴, \hat{\sigma}_۲ = ۰/۲۳۷۱۴۷$$
$$\alpha_۲ = ۲/۳۷۶۸, \hat{\delta}_۲ = ۰/۰۳۸۲ \qquad (۳.۸)$$

که تابع چگالی پسین حاصل از بیماران بفرم $\frac{\phi(t+x+۰/۴۲۰۷)}{\psi(x+۰/۴۲۰۷)}$ و تابع امید ریاضی و واریانس آن نیز برابر است با:

$$M_۲(x) = \frac{\phi(x+۰/۴۲۰۷)}{\psi(x+۰/۴۲۰۷)} - (x+۰/۴۲۰۷)$$

$$V_۲(x) = ۱ + \frac{(x+۰/۴۲۰۷)^۲}{۴} - \left\{\frac{\phi(x+۰/۴۲۰۷)}{\psi(x+۰/۴۲۰۷)} - \frac{(x+۰/۴۲۰۷)}{۲}\right\}^۲ \qquad (۴.۸)$$

برای آزمون برابری $\mu_1 = \mu_2$ ابتدا آزمون برابری واریانس‌ها را انجام می دهیم

$$s_1^2 = \frac{n_1}{n_1 - 1}\hat{\sigma}_1^2 = \frac{۱۹۱}{۱۹۰}(۰/۳۵۴)^2 = ۰/۱۲۵۹۷$$

$$s_2^2 = \frac{n_2}{n_2 - 1}\hat{\sigma}_2^2 = \frac{۴۷۷}{۴۷۶}(۰/۲۳۷۱۴۷)^2 = ۰/۰۵۶۳۵$$

$$\frac{s_2^2}{s_1^2} = ۰/۴۴۷۳۸ \qquad (۵.۸)$$

با توجه به جدول توزیع فیشر برابری واریانس دو جامعه را در سطح ۰/۰۵ می پذیریم.

با فرض واریانس‌های برابر و مجهول آزمون برابری میانگینها را انجام می‌دهیم.

$$s_p^2 = \frac{(n_1 - 1)s_1 + (n_2 - 1)s_2}{n_1 + n_2 - 2} = ۰/۰۷۶۲۱$$

آماره آزمون مربوطه با فرض برابری میانگین‌ها(درستی فرض صفر) برابر است با

$$T = \frac{۶/۹۶۶ - ۵/۸۲۴}{\sqrt{۰/۰۷۶۲۱}\sqrt{\frac{۱}{۴۷۷} + \frac{۱}{۱۹۱}}} = ۴۸/۳۱۱ \qquad (۶.۸)$$

که دلیلی برای پذیرش فرض صفر یعنی برابری میانگین‌ها وجود ندارد.

**توجه ۸. ۱:** از بحث فوق نتیجه می‌گیریم که متوسط کاهش $R(T)$ (لگاریتم cd4 در اولین آزمایش) در بیماران شهر نیویورک بیشتر از بیماران شهر کرمانشاه است.

**توجه ۸. ۲:** نمودار ۳ نشان‌دهنده متوسط دوره پنهانی(واحد اندازه‌گیری ماه است) ویروس ایدز در مقابل تعداد cd4 شمارش شده در اولین ملاقات دو شهر کرمانشاه و نیویورک می‌باشد. بوضوح می توان دید که متوسط دوره پنهانی در شهر کرمانشاه بسیار کمتر است و می‌توان گفت پس از ورود ویروس ایدز علائم بیماری در بیماران کرمانشاه زودتر از بیماران نیویورک مشخص می شود.

**توجه ۸. ۳:** چگالی‌های پسین دو شهر در نمودارهای ۴، ۵ و ۶ با یکدیگر مقایسه می‌شوند این نمودارها بترتیب مربوط به مشاهده تعداد cd4 ۲۰۰، ۶۰۰ و ۱۰۰۰ در اولین آزمایشی است که منتهی به تشخیص بیماری ایدز می‌شود

**توجه ۸. ۴:** در جدول ۱ متوسط دوره پنهانی ویروس ایدز و صدک ۹۵ توزیع پسین برای تعدادی از مقادیر cd4ها آمده است. صدکها براحتی از جدول توزیع نرمال قابل محاسبه می‌باشند.

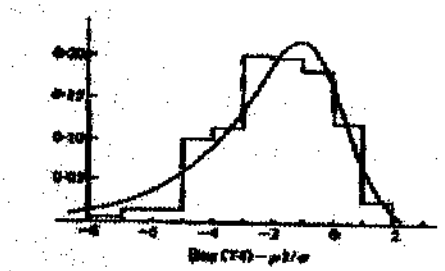
**نتیجه ۸. ۱:** می توان گفت متوسط زمان افت میزان cd4 در بیماران شهر کرمانشاه‌بسیار کمتر از بیماران نیویورک می‌باشد.

نقطه‌ای که $\alpha$ درصد احتمال بالای آن باشد همانند $t_\alpha$ نقطه‌ایست که در رابطه زیر صدق کند:

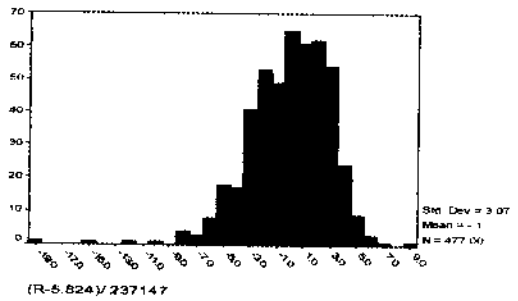
$$\psi(t_\alpha + x + n^{-1}) = \alpha \times \psi(x + m^{-1}).$$

**نتیجه ۸. ۲:** با توجه به نمودار ۴(مربوط به تعداد cd4 ۲۰۰) می‌توان بیان کرد جائیکه چگالی پسین مربوط به داده‌های کرمانشاه تقریباً صفر است( $t \geq ۳۰$ ) چگالی پسین داده‌های نیویورک بیشترین مقادیر خود را می‌گیرند. بدین معنی که با مشاهده تعداد cd4 ۲۰۰ در بیماران کرمانشاه مدت دوره پنهانی کمتر مساوی ۳۰ ماه است ولی در بیماران شهر نیویورک این زمان بیشتر از ۳۰ ماه است.

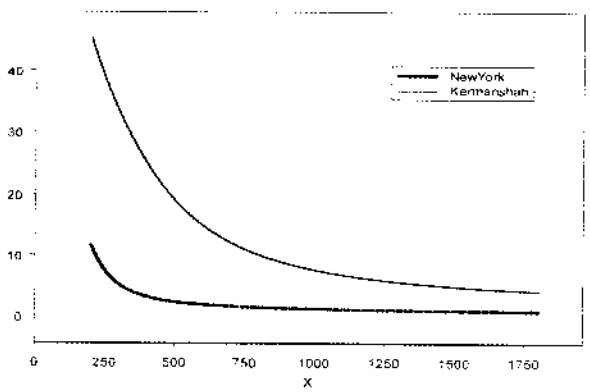
**نتیجه ۸. ۳ :** اگر بیمار مبتلا به ایدز شهر کرمانشاه که از طریق تزریق دچار این بیماری شده است تعداد cd4 خون این شخص در اولین زمان تشخیص بیماری ۲۰۰۰ باشد با اطمینان ۹۵ درصد می‌توان بیان کرد که حداکثر ۵۴ روز از زمان عفونت این بیمار سپری می‌شود.



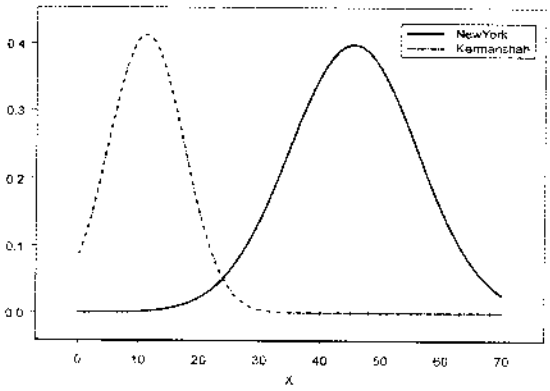
**نمودار ۱: هیستوگرام مقادیر $(R - \mu)/\sigma$ در مقایسه با چگالی (٦,٥)**



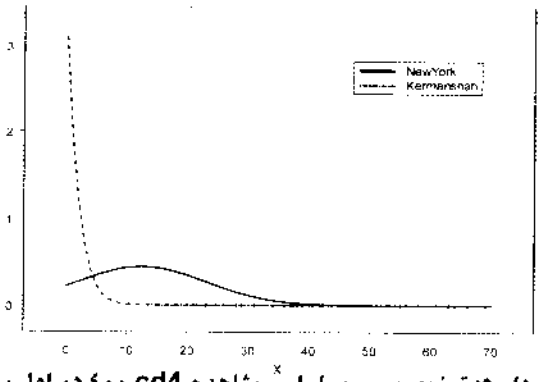

**نمودار۲: هیستوگرام مقادیر $(R - \mu)/\sigma$ در شهر کرمانشاه**



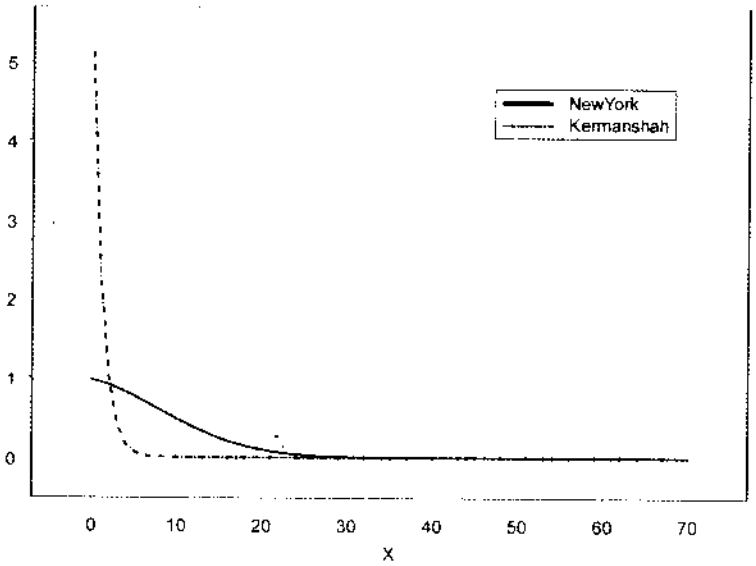

نمودار ۳: متوسط دوره پنهانی ویروس ایدز در بیماران شهر کرمانشاه و نیویورک




نمودار ٤: توزیع پسین بر اساس مشاهده cd4 ۲۰۰ در اولین ملاقات




نمودار ٥: توزیع پسین بر اساس مشاهده cd4 ٤۰۰ در اولین ملاقات




نمودار۶: توزیع پسین بر اساس مشاهده cd4 ۱۰۰۰ در اولین ملاقات

جدول ۱: میا نگین و صدک ۹۵م تابع توزیع پسین

| تعداد T4 | متوسط زمان (ماه) | | صدک نود و پنجم(ماه) | |
|---|---|---|---|---|
| | نیویورک | کرمانشاه | نیویورک | کرمانشاه |
| ۲۰۰ | ۴۵/۳۳ | ۱۱/۶۶ | ۶۲/۷۱ | ۲۱/۴۷ |
| ۴۰۰ | ۲۴/۹۲ | ۳/۱۱ | ۴۲/۰۷ | ۸/۴۳ |
| ۶۰۰ | ۱۴/۹۰ | ۱/۸۳ | ۳۰/۵۵ | ۵/۲۷ |
| ۸۰۰ | ۱۰/۰۴ | ۱/۳۹ | ۲۳/۴۸ | ۴/۰۶ |
| ۱۰۰۰ | ۷/۵۲ | ۱/۱۶ | ۱۹/۰۱ | ۳/۴۲ |
| ۱۲۰۰ | ۶/۰۶ | ۱/۰۲ | ۱۶/۰۵ | ۳/۰۲ |
| ۱۴۰۰ | ۵/۱۴ | ۰/۹۳ | ۱۳/۹۹ | ۲/۷۴ |
| ۱۶۰۰ | ۴/۵۱ | ۰/۸۶ | ۱۲/۵۰ | ۲/۵۴ |
| ۱۸۰۰ | ۴/۰۵ | ۰/۸۳ | ۱۱/۳۷ | ۲/۳۹ |
| ۲۰۰۰ | ۳/۷۰ | ۰ | ۱۰/۴۹ | ۱/۸۴ |

- نویسنده از بخش اطلاعات و آمار شبکه بهداشت و درمان شهرستان کرمانشاه، که هماهنگی و همکاری لازم را جهت اخذ این داده ها داشته اند تشکر می کند.

## مراجع

[1] Cramer, H., 1946, *Mathematical Methods of Statistics,* Princeton University Press

[2] De Gruttola, V. & Lange, N., 1989, Statistical modeling of T4 counts between infection and AIDS diagnosis, *Bull. Int. Statist. Inst.*, 47 Session, Paris, pp.101-115

[3] Des Jarlais, D. C., Friedman, S. R., Marmor, M., Cohen, H., Mildavan, D., U., El-Sadr, W., Spira, T. J., Garber, J., Beatrices, S, T., Abdul- Quader, A. S.& Sotheran, J. L.,1987, Development of AIDS, HIV seroconversion, and potentialco-factors for T4 cell loss in a cohort of intravenous drug users. *AIDS,* 1, 105-111.

[4] Marmor, M., Des Jarlais, D. C., Cohen, H., Friedman, S. R., Beatrices, S. T., Dubin,N., El-Sadr, W., Mildavan, D., Yancovitz, S., Mathur, U., & Holzman, R.1987. Risk factor for infection with human immunodeficiency virus among intravenous drug abusers in New York City, *AIDS,* 1, 39-44.

[5] Rossi, C. et al,1989, A Stochastic approach to modeling incubation period in HIV Infection, *Bull. Int. Statist. Inst.*, 46 Session, Paris.

[6] Simeon M. Berman,1990, stochastic model for the distribution of HIV latency time based on T4 count, Biometrica, 77, vol 4, pp. 733-741.

[7] Yang, T. & Dubin, N., 1990, Nonparametric density estimation for immunologic Measurements, To appear.