دانشگاه پیام نور

**دانشکده علوم پایه**

**پایان نامه**

**برای دریافت مدرک کارشناسی ارشد**

**رشته آمار ریاضی**

**گروه آمار**

**عنوان پایان نامه :**

# استنباط بیزی تغییرات در مدل های رگرسیونی

**عصمت هاشمی**

**استاد راهنما :**

**دکتر نرگس عباسی**

**مهر ۱۳۹۲**

بسم الله الرحمن الرحيم

دانشگاه پیام نور

**دانشکده علوم پایه**

**مرکز شیراز**

**پایان نامه**

**برای دریافت مدرک کارشناسی ارشد**

**رشته آمار ریاضی**

**گروه آمار**

# عنوان پایان نامه:

# استنباط بیزی تغییرات در مدل رگرسیونی

**عصمت هاشمی**

**استاد راهنما**:

**دکتر نرگس عباسی**

**مهر ۱۳۹۲**

جمهوری اسلامی ایران
وزارت علوم، تحقیقات و فناوری
دانشگاه پیام نور استان فارس

دانشگاه پیام نور شیراز

باسمه تعالی

تاریخ : ۹۲/۰۷/۲۰
شماره : ۰۵/۱۶۲۷۶
پیوست :

## صورتجلسه دفاع از پایان‌نامه دوره کارشناسی ارشد

جلسه دفاع از پایان‌نامه دوره کارشناسی ارشد **خانم عصمت هاشمی** دانشجوی رشته **آمار ریاضی** به شماره دانشجویی **۹۱۷۴۱۴۱۶۵** با عنوان:

**" استنباط بیزی تغییرات در مدل‌های رگرسیونی"**

با حضور هیأت داوران در روز **شنبه** مورخ **۱۳۹۲/۰۷/۲۰** ساعت **۱۶:۳۰** در محل ساختمان **غدیر دانشگاه پیام نور شیراز** برگزار شد و هیأت داوران پس از بررسی، پایان‌نامه مذکور را شایسته نمره به عدد ۱۸.۷ به حروف هیجده و هفت دهم با درجه عالی تشخیص داد.

| ردیف | نام و نام خانوادگی | هیات داوران | مرتبه دانشگاهی | دانشگاه | امضاء |
|---|---|---|---|---|---|
| ۱ | دکتر نرگس عباسی | راهنما | دانشیار | پیام‌نور شیراز | |
| ۲ | عبدالرضا بازرگان‌لاری | داور | استادیار | شیراز | |
| ۳ | امیر اکبری | نماینده تحصیلات تکمیلی | مربی | پیام‌نور شیراز | |

**رئیس اداره تحصیلات تکمیلی**

دانشگاه پیام نور مرکز شیراز
معاونت آموزشی و تحصیلات تکمیلی

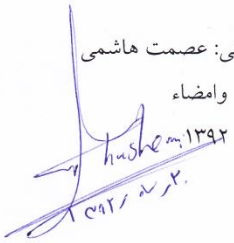
شیراز- شهرک گلستان.بلوار دهخدا
قبل از نمایشگاه بین المللی
تلفن : ۶۲۲۲۲۵۵ - ۰۷۱۱
دورنگار : ۶۲۲۲۲۴۹ - ۰۷۱۱
صندوق پستی: ۱۳۶۸ - ۷۱۹۵۵
www.spnu.ac.ir
Email : admin@spnu.ac.ir

گواهی اصالت، نشروحقوق مادی و معنوی اثر:

اینجانب عصمت هاشمی دانشجوی ورودی ۱۳۹۱ مقطع کارشناسی ارشد رشته آمار ریاضی گواهی می نمایم چنانکه در پایان نامه خود از فکر،ایده و نوشته دیگری بهره گرفته ام با نقل قول مستقیم یا غیر مستقیم منبع و ماخذآن را نیز در جای مناسب ذکر کرده ام بدیهی است مسئولیت تمامی مطالبی که نقل قول دیگران نباشد بر عهده خویش میدانم و جوابگوی آن خواهم بود.
دانشجو تائید مینماید که نتایج مندرج در این پایان نامه نتیجه تحقیقات خودش میباشد در صورت استفاده از نتایج دیگران مرجع آن را ذکر نموده‌ام.

نام ونام خانوادگی: عصمت هاشمی
تاریخ و امضاء

اینجانب عصمت هاشمی دانشجوی ورودی سال ۱۳۹۱ مقطع کارشناسی ارشد رشته آمار ریاضی گواهی مینمایم چنانچه براساس پایان نامه خود اقدام به انتشار مقاله،کتاب و...نمایم،ضمن مطلع نمودن استاد راهنما با نظر ایشان نسبت به نشر مقاله،کتاب و...به صورت مشترک و باذکر نام استاد راهنما مبادرت نمایم.
کلیه حقوق مادی مرتبط از نتایج مطالعات آزمایش و نوآوری ناشی از تحقیق موضوع پایان نامه متعلق به دانشگاه پیام نور می‌باشد.



نام و نام خانوادگی: عصمت هاشمی
تاریخ وامضاء
مهر ۱۳۹۲

تقدیم به:

حضرت مهدی صاحب الزمان (عج):

چه انتظار عجیبی!

تو بین منتظران هم، عزیز من، چه غریبی!

عجیب‌تر که چه آسان، نبودنت شده عادت، چه بی‌خیال نشستیم

نه کوششی، نه وفایی ...

فقط نشستیم و گفتیم: خدا کند که بیایی ...!

## سپاسگزاری

سپاس خدایی را که هرچه دارم از اوست

به امید آنکه توفیق یابم جز خدمت به خلق او نکوشم

به مصداق «من لم یشکر المخلوق لم یشکر الخالق» بسی شایسته است از استاد فرهیخته و فرزانه سرکار خانم دکتر نرگس عباسی که با کرامتی چون خورشید سرزمین دل را روشنی بخشید و گلشن سرای علم و دانش را با راهنمایی‌های کارساز و سازنده بارور ساختند تقدیر و تشکر کنم هم چنین از جناب اقای دکتر عبدالرضا بازرگان لاری که زحمت داوری این پایان نامه را برعهده داشتند تقدیر و تشکر کنم.

تقدیم به پدر و مادرم که از انگشتانشان صلابت,از رفتارشان محبت واز صبرشان ایستادگی را آموختم که نه می‌توانم موهایشان را که در راه عزت من سفید شد سیاه کنم و نه برای دستهای پینه بستان که ثمره تلاش برای افتخار من است مرهمی دارم پس توفیق ده که هر لحظه شکرگزارشان باشم و ثانیه‌های عمرم را در عصای دست بودنشان بگذرانیم.

# چکیده

در این پایان نامه روشی موثر در برآوردیابی پارامترهای مجهول در مـدلهای خطـی و شـبه خطـی در تحقیقات علوم پزشکی و انسانی مورد استفاده قرار گرفت که روشهای ارائه شده در ایـن پایـان نامـه منجر به محاسبه ی توزیع پسین خواهد شد.هدف ارائه یک الگوریتم جهت محاسبه پارامترها در کلیه مسائل استنباطی است که کاربردهای این تحقیق ارائه رویکردی جدیـد در مـدل بنـدی و اسـتفاده از مباحث رگرسیون در تحقیقات با داده‌های واقعی با استفاده از نظریه بیز است.

در مدلهای رگرسیون کلاسیک به دنبال برآورد یک نقطه از بردارهای پارامتر w هسـتیم. در مقابـل در یک رویکرد بیزی عدم قطعیت درw را از طریق یک توزیع احتمال p(w) مشخص می‌کنیم.

مشاهدات از نقاط داده‌ها توسط قضیه بیـز بـا داده‌هـایی کـه از طریـق تـابع احتمـال در درسـتنمایی واسطه‌اند تغییر میاد.

**واژگان کلیدی:**مینی ماکس، قابلیت قبول، استنباط بیزی تغییرات، رگرسیون لجستیک، رگرسـیون خطی، رگرسیون بیزی، ماشین بردار پشتیبانی

**فهرست مطالب:**

**عنوان** **صفحه**


مقدمه..... ۱

فصل اول: تئوری بیزی..... ۲

۱-۱ – مقدمه..... ۳

۱-۲- مفاهیم بیزی..... ٤

۱-۲-۱- تابع زیان..... ٤

۱-۲-۲- مینیماکس ..... ٦

۱-۲-۳- تعریف کار مینیماکس..... ٦

۱-۲-٤- تعریف کار بیزی..... ۷

۱-۲-٥- مخاطره ..... ۹

۱-۲-٦- تابع مخاطره..... ۱۰

۱-۲-۷- تعریف قاعده تصمیم مینیماکس..... ۱۰

۱-۳- قاعده ی تصمیم بیزی..... ۱۱

۱-۳-۱- قواعد تصمیم تصادفی..... ۱۲

۱-۳-۲- پارامترهای مزاحم..... ۱۲

۱-۳-۳- تابع زیان برای قواعد تصمیم تصادفی..... ۱۳

۱-٤- طبقه بندی و آزمون فرضیه..... ۱۳

۱-٤-۱- آزمون فرضیه..... ۱۳

۱-٤-۲- روش آزمون فرضیه..... ۱٤

۱-٤-۳- طبقه بندی..... ۱۹

۱-٤-٤- برآورد..... ۲۰

۱-۵- قابلیت قبول ..... ۲۳
۱-۵-۱- کمال ..... ۲٤
۱-۵-۲- قابلیت قبول و کمال ..... ۲۵
۱-۵-۳- قابلیت قبول و مینیماکس ..... ۲٦
۱-۵-٤- قابلیت قبول وبیز ..... ۲۷
۱-۵-۵- قاعده ی بیز تعمیم یافته ..... ۲۸
۱-۵-٦ -کامل بودن ..... ۳٤
۱-۵-۷- بسندگی و نابرابری رائو بلکول ..... ۳٦
فصل دوم: استنباط بیزی تغییرات ..... ۳۹
۲-۱- مقدمه ..... ٤۰
۲-۲- یک نمونه ساده ..... ٤۱
۲-۲-۱- فرض استقلال توزیع‌ها ..... ٤۲
۲-۲-۲- بدست اوردن عامل $q(\mu)$ ..... ٤۳
۲-۲-۳- بدست اوردن فاکتور $\mathrm{q}(\tau)$ ..... ٤۵
۲-۲-۴- الگوریتم محاسبه ی پارامترهای بهینه سازی ..... ۴۶
۲-۳- یک نمونه پیچیده‌تر(مدل مخلوط گوسی) ..... ٤۹
فصل سوم : رگرسیون لجستیک ..... ۵۵
۳- ۱-رگرسیون لجستیک ..... ۵٦
۳-۲ -استنتاج بیزی برای پارامترهای مدل رگرسیون لجستیک ..... ٦٤
۳-۳- مدل رگرسیون خطی کلاسیک ..... ٦٦
۳-۳-۱ - مدل رگرسیون خطی ..... ٦۹
۳-۳-۲- مثالها ..... ۷٦

۳-٤- رگرسیون لجستیک بیزی ..... ۷۹
۳-٥- استنتاج بیزی تغییرات ..... ۸۰
۳-٦- چگالی پیش بینی ..... ۸٥
۳-۷- اجرا ..... ۹۰
۳-۸ -توزیع پسین خمیده ایجاد شده ..... ۹۲
۳-۹- چگالی پیش بینی شده ..... ۹٤
۳-۱۰- محدوده تغییرات ..... ۹٤
**فصل چهارم** : لجستیک بیزی و طبقه بندی ..... ۹٥
٤-۱ -رگرسیون بیزی و طبقه بندی ..... ۹٦
٤-۲ -تنظیم ..... ۹۷
٤-۳- مدلهای احتمالی ..... ۹۸
٤-٤ -رگرسیون بیزی ..... ۹۹
٤-٥ -ماشین بردار پشتیبانی ..... ۱۰۱
٤-٦ -مدلهای خاص ..... ۱۰۲
٤-۷ -پیشین موثر ..... ۱۰٤
٤-۸ -پیش بینی ..... ۱۰٦
٤-۹ -طبقه بندی ..... ۱۰۹
٤-۱۰ -دادههای مصنوعی گویا ..... ۱۱۱
٤-۱۱ -نتیجه معیار ..... ۱۱۳
منابع ..... ۱۱٥

**فهرست شکل‌ها:**

**عنوان** **صفحه**

شکل (۲-۱) ..... ٤٩

شکل (٤-۱) ..... ۱۱۲

## مقدمه

در فصل اول به معرفی مفاهیم اولیه در نظریه بیز,تابع ریسک و اصل بهینگی می‌پردازد، در پایان آزمون فرض‌ها را مطرح و ارتباط آن با مباحث بیز را بیان می‌کنیم و سپس قابلیت قبول و قضایای مربوط به آن را بحث خواهیم کرد. فصل دوم به معرفی استنباط بیزی تغییراتی می‌پردازد که ازجمله روش‌های یادگیری بیزی است که برای تقریب جواب با استفاده از یک سری فرض‌های استقلال برای توزیع پسین بکار می‌رود هدف در این فصل ارائه یک الگوریتم جهت محاسبه‌ی پارامترها در مسائل استنباطی است. در فصل سوم به معرفی رگرسیون لجستیک و خطی می‌پردازیم سپس استنباط آماری را برای هر کدام به کار می‌بریم این مدل را می‌توان به عنوان مدل خطی تعمیم یافته‌ای که از تابع لوجیت به عنوان تابع پیوسته استفاده می‌کند حساب آورد، سپس مدل رگرسیونی خطی کلاسیک را بیان و در پایان هم رگرسیون لجستیک بیزی را بیان می‌کنیم. در فصل چهارم برخی از مفاهیم اساسی رگرسیون طبقه‌بندی از دیدگاه‌های بیزی مشخص شده است جزئیات مدل بیزی خاص را به نام ماشین بردار پشتیبانی که منجر به راه حل‌های پراکنده و داشتن خواص تعمیم عالی است مورد بحث قرار می‌دهیم و در پایان هم پیش‌بینی را براساس توزیع پسین با حداکثر رساندن یک داده‌ی جدید انجام می‌دهیم سپس تابع کرنلی برای تعریف داده‌های مصنوعی گویا استفاده می‌کنیم.

# فصل اول

# تئوری بیزی

## ۱–۱ مقدمه

در این فصل به معرفی مفاهیم اولیه در نظریه بیز، تابع ریسک و اصل بهینگی می‌پردازیم. در پایان آزمون فرض‌ها را مطرح و ارتباط آن در مباحث بیز را بیان می‌کنیم.

تابع تصمیم: این فن بر فن معماری تئوری آزمون‌های آماری می‌پردازد. یک فرمول قصد در تهیه یک بنیاد منطقی که یاد می‌گیریم از داده‌ها را دارد. این نظریه یک نظریه وسیع است و گسترش مفاهیمی از چند دهه و چند دیدگاه متفاوت می‌باشد. در اواخر ۱۸۸۰ یک دوره‌ای که ماشین آلات نقش برجسته‌ای در دانش اقتصادی و اجتماعی داشته، شاخه‌ای برتر ازمدل‌های احتمالی بود، قاعده و اساس این کاربرد پذیری ریشه در قسمت‌های از موضوع از تئوری مشاهده شده دارد که به طور واضح در اندازه‌گیری کمیت‌های فیزیکی بحث کرده‌اند.

تاریخچه و بیوگرافی جزئیات در ویز[1] است. او در مقاله ۱۹۴۹ روی تابع توزیع آماری والد[2] یک چهارچوب متحدی برای به وجود آوردن تئوری آماری که برروی رفتارهای استنتاج آماری به عنوان یک نمونه خاص از تئوری بازی‌ها مبتنی است پیشنهاد داد. تئوری ریاضیات از بازی‌ها که پایه و اساس برای تئوری‌های اقتصادی فراهم می‌کند توسط نیمن[3] و مورگنسترن[4] (۱۹٤٤) پیشنهاد داده شده است که شامل رفتار بحث‌های تئوری متحد و یکتاست.

والد استنتاج آماری را به شکل یک بازی دو نفره با حاصل جمع صفر نشان داد یکی از آنها طبیعی و دیگری آماری است.

[1] Weiss
[2] Wald
[3] Neyman
[4] Morgensten

انتخاب طبیعی توزیع احتمالات برای آزمایش است که به وسیله‌ی آمار مشاهده کردیم این بخش از توزیع‌های آماری به اصل مینیماکس مبتنی است که در این قسمت بحث خواهیم کرد. اصل مینیماکس یک استراتژی محافظه‌کارانه توزیع آماری را فراهم می‌کند این اصل در تئوری بازی‌ها به توزیع پیشین وقتی که ناشناخته است احتیاج ندارد.

که این برای تشخیص دادن دو جنبه از همکاری والد مفید است یکی از آنها از مسائل توزیع آماری دیگری معقول بودن اصل ادعا کردن برای حل آن می‌باشد، تفصیل‌های معماری خودش برای بدست آوردن توزیع آماری تحت اصل سودی که انتظار داریم بخوبی معطوف شده است.

تحقیقات نشان داده‌اند که تابع توزیع آماری نتایج بسیاری در این وجود شامل فرگوسن[1] و روبرت[2] است که از این مفاهیم و ابزارها برای نتایج بیزی می‌توانیم استفاده کنیم.

## ۱–۲ مفاهیم بیزی

### ۱–۲–۱ تابع زیان

بحث در این بخش شامل دو مجموعه از کارها است که نتیجه بستگی به حالت‌های ناشناخته والد یا حالت‌های از طبیعت دارد که یک مجموعه از کارها $\mathcal{A}$ و عضوهای آن مجموعه $a$ نامیده می‌شود یک مجموعه از حالت‌ها $\Theta$ با عضوهای $\theta$ نامیده می‌شود، برای انتخاب میان کارها هدف ارزیابی مقدار آنها است. این ارزیابی باید تابعی از $a$ و $\theta$ باشد.

بنابراین با $U(a,\theta)$ کار می‌کنیم که آماردانان از عضو پیوسته $(a,\theta) \in (\mathcal{A},\Theta)$ و تعریف یک تابع نرمال استفاده می‌کنند.

[1] Fergusen
[2] Robert

در نظریه والد و در بسیاری از نظریه‌های تصمیم آماری باخت در اثر انتخاب یک عمل زمانی که وضعیت درست از ماهیت $\theta$ باشد متناسب با باخت‌هایی است که در اثر اقدامات دیگر به وجود می‌آید در یکی از مثال‌های که والد نوشته است وزن تابع $L(a,\theta)$ برای هر نقطه $\theta$ عضو $\Theta$ و هر عضو $a$ از $\mathcal{A}$ یک مقدار واقعی غیرمنفی تعریف شده است که ارتباط مهم توسط پذیرش $\mathcal{A}$ وقتی که $\theta$ درست باشد را بیان می‌کند. اگر $\theta$ در $L(a,\theta)$ محاط باشد البته برابر صفر است.

اگر تابع مطلوبیت مشخص باشد می‌توانیم سود را همانند زیان $(a,\theta) \in (A,\Theta)$ بیان کنیم چنانکه

(۱-۱)

$$u\big(a(\theta)\big) = -L(\theta,a)$$

با این حال, اگر کسی از توابع سودمندی داده شده شروع کند هیچ تضمینی وجود ندارد که برای یک $\theta$ مفروض باید عملیاتی با باخت صفر وجود داشته باشد.

این شرایط احتیاج به اطلاع بیشتر سود که به تابع زیان $L(a,\theta)$برمیگردد, دارد.

(۱-۲)

$$L(a,\theta) = \big(\mathrm{L_u}(a,\theta) - \ inf_{a\in A} L_u(\theta,a)$$

می توانیم تابع زیان را مستقیم از مطلوبیت مشخص کنیم.

(۱-۳)

$$L(a,\theta) = \sup_{a'(\theta)} (u(a'(\theta)) - u(a(\theta))$$

قبل از توضیح دادن دلایل برای این, احتیاج به معرفی اصل مینیماکس و اصل مطلوبیت انتظار رفته در ۲ قسمت بعد داریم.

## ۱-۲-۲ مینیماکس

اصل مینیماکس از انتخاب تئوری آزمون آماری روی تئوری بازی‌ها مبتنی است و کـاربرد مهمـی در تئوری بازی‌ها و بهینه سازی دارد. فرض می‌کنیم که تابع زیان نشان دهنده سـاختار سـود هـم بـرای آمارگران و هم مخالف (ذات) باشد. ذات طبیعی اولین را انتخاب می‌کند و بنابراین بهترین اسـتراتژی برای آمارگران این است که بدترین چیز را فرض کنند و عملی را انتخاب کنند که حـداکثر امتیـاز از دست دادن را به حداقل برسانند.

## ۱-۲-۳ تعریف کار مینیماکس

$a^{\mathrm{M}}$ مینیماکس است اگر

(۱-٤)

$$a^{\mathrm{M}} = \arg\min \max_{\theta} \mathrm{L}(\theta, a)$$

## ۱–۲–۴ تعریف کار بیزی

یک کار $a^*$ بیزی است اگر

(۱–۵)

$$a^* = arg\min \int_\theta L(\theta, a)\pi(\theta)\, d\theta$$

که نشان می‌دهیم اگر $\pi$ تابع چگالی پیشین باشد آنگاه:

(۱–۶)

$$\mathcal{L}_\pi(a) = \int L(\theta, a)\ \pi(\theta)\, d\theta$$

فرمول‌های بهینه سازی تفاوت بین(5-1) و (6-1) شبیه به این است که عملگر انتظار جایگزین عملگر مینیماکس است هر ۲ عملگر یک راهی برای بررسی حالتهای نامعلوم جهان تهیه می‌کنند در این قسمت ارتباط بین بیز و تصمیم مینیماکس را که در این بخش استفاده می‌کنیم معرفی می‌شود.

که این مثال انتخاب یک شرکت مخابرات برای ارتباط‌های بین المللی است که A,B شرکت‌های مخابراتی هستند به وسیله بیز نشان میدهیم که خدمات شرکت A ارزانتر است.

تابع زیان فرمول‌های زیر در نظر گرفته شده است

$$L(\theta, a) = 2\theta \quad \theta\epsilon[0,1]\,, \quad L(\theta, B) = 1$$

در این جا مقدار ۱ نشان دهنده تفاوت میان هزینه اشتراک شرکت B و هزینه اشتراک شرکت A است. همکار شما به این مقدار یک تابع خطی از تعداد تماسهای از دست رفته اضافه می‌کند بنابراین اگر او شرکت B را انتخاب کند تابع زیان باید شناخته شده باشد اگر شرکت A را انتخاب کند تابع زیان بستگی به این دارد که در قسمتی از زمان او برای تلفن‌های بین المللی شکست خواهد

خورد.اگرθرا ۰/۲۵ در نظر بگیریم در این صورت تابع زیان ۰/۵ است عمـل مینیمـاکس را مـی‌تـوان بدون هیچ گونه ورودی بیشتر محاسبه کرد و برای انتخاب شرکت B می‌باشد زیرا:

$$sup_\theta L(\theta, A) = 2 > 1 = sup_\theta L(\theta, B)$$

اندکی محافظه کارانه به نظر می‌رسد که شرکت A اگر بیش از نیمی از تماس‌هایش را از دست دهـد برBبرتری دارد براساس یک نظر سنجی از گزارش‌های مصرف کننـدگان انحـراف معیـار ۰/۱۴۸۷ و میانگین ۰/۰۴۷۶ اندازه گیری شده است

بنابراین تصمیم می‌گیریم که توزیع بتا یک انتخاب معقول برای توزیع پیشین رویθ باشد با اسـتفاده از:

$$\frac{\alpha\beta}{\alpha+\beta} = \bar{x} \text{و} \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = S^2$$

پارامترهای اضافی $\alpha_0 = 0/05$ و$\beta_0 = 1/00$ می‌باشد و تابع چگالی احتمالات پیشین هست:

$$\pi(\theta) = 0/05\theta^{\cdot/95}$$

متوسط تابع زیان پیشین برابر است با

$$\int_0^1 L(\theta, a)\ \pi(\theta)\, d\theta = \begin{cases} \int_0^1 2\theta\ \pi(\theta)\, d\theta = 2E_\theta[\theta]\ \ if\ a = A \\ \int_0^1 1\ \pi(\theta)\, d\theta = 1 \qquad if\ a = B \end{cases}$$

بنابراین $2E_\theta[\theta] = 2 * 0/05\ /(1 + 0/05) \quad = 0/095$ که از ۱ کمتر است که کارهای بیزی برای تقاضا برای شرکت A است. بنابراین بیز و مینیمـاکس در ایـن مثـال ۲ نتیجـه متفـاوت را می‌دهد که این تفاوت به این واقعیت که θناشناخته است بر می‌گردد اگر به جای بررسی گزارش‌های مصرف کنندگان شما $\pi(\theta) = 1$ را انتخاب کنید یک پیش فرض یکسان ممکن است نشان دهنـده