

تحلیل منحنی‌های ROC برای مقایسه تست‌های تشخیصی پزشکی

سیدمهدی سادات‌هاشمی^۱ (Ph.D)، راهب قربانی^۱ (Ph.D)، بهروز کاوه‌بی^۲ (Ph.D)

۱- دانشگاه علوم پزشکی سمنان، دانشکده پزشکی، گروه پزشکی اجتماعی

۲- سازمان سنجش آموزش کشور

چکیده

سابقه و هدف: با مسائل تشخیصی هنگامی مواجه می‌شویم که هدف از یک آزمایش یا مدل، پیش‌بینی و انتساب افراد یا اشیاء (نمونه‌ها) به گروه‌های معین و خاصی باشد. به عنوان مثال در پزشکی برای تشخیص دیابت و یا سرطان (که در حیطه پیش‌گیری سطح دوم قرار می‌گیرند)، از ملاک‌ها و شاخص‌های مختلفی استفاده می‌شود. ساده‌ترین حالت هنگامی است که هدف تخصیص نمونه‌ها به یکی از دو گروه ممکن (سالم یا بیمار) باشد. در این صورت صفتی در مقیاس پیوسته و یا رتبه‌ای بر روی نمونه‌ای از افراد سالم و بیمار اندازه‌گیری می‌شود و سپس با انتخاب یک نقطه برش مناسب برحسب میزان حساسیت و ویژگی تست در آن نقطه و نیز مقدار تابع زیان اهمیت تشخیصی، افراد سالم و بیمار مشخص می‌گردند. یک شاخص ارزیابی مناسب و تک مقداری برای تست در این حالت، عبارت است از آماره A ، مساحت زیر منحنی (Receiver operating characteristic, ROC). از آن‌جا که ملاک‌ها و شاخص‌ها بر روی نمونه واحدی اندازه‌گیری می‌شوند، لذا یک هم‌بستگی طبیعی بین A های متناظر با تست‌های مختلف وجود خواهد داشت که هرگاه هدف، مقایسه تست‌ها و انتخاب بهترین تست از بین بقیه باشد، صرف نظر کردن از این هم‌بستگی می‌تواند منجر به نتایج مخدوش و گمراه‌کننده شود.

مواد و روش‌ها: در این مقاله ابتدا نحوه محاسبه آماره‌های A و ماتریس وارینانس-کوواریانس آن‌ها را به همراه آزمون مناسبی برای مقایسه A های هم‌بسته تشریح نموده و با استفاده از داده‌های شبیه‌سازی شده نشان داده‌ایم که آزمون چگونه تحت تأثیر منظور نمودن هم‌بستگی‌ها قرار می‌گیرد. برای این کار برنامه مناسبی را با استفاده از زبان برنامه نویسی Delphi5 تهیه نموده‌ایم.

یافته‌ها: بر اساس داده‌های شبیه‌سازی شده برای مقایسه دو تست، به‌دست آوردیم:

$A_1 = 0.66$ ، $SE(A_1) = 0.054$ ، $A_2 = 0.49$ ، $SE(A_2) = 0.060$ ، لذا $Z = 2.1$ منجر به رد فرض برابری A_1 و A_2 در سطح معنی‌داری $\alpha = 0.05$ شده است. چنان‌چه هم‌بستگی A_1 و A_2 در نظر گرفته شود، $Z = 1.92$ که منجر به پذیرش فرض برابری A_1 و A_2 شده است.

نتیجه‌گیری: همان‌طور که از نتایج مشاهده می‌شود، چنان‌چه از کوواریانس بین A_1 و A_2 صرف‌نظر شود، نتیجه آزمون می‌تواند گمراه‌کننده بوده و منجر به انتخاب تست‌های نامناسب شود.

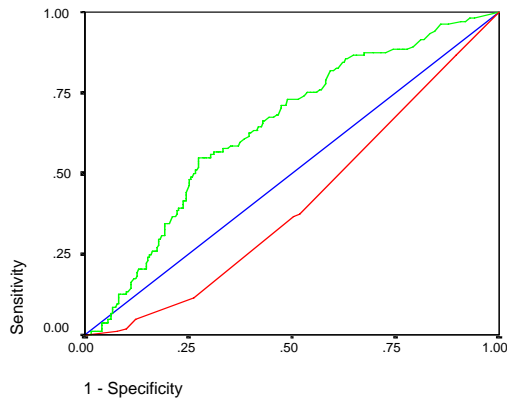
واژه‌های کلیدی: منحنی‌های ROC، حساسیت، ویژگی، تست تشخیصی، هم‌بستگی

مقدمه

در علم پزشکی هدف از بررسی بالینی و یا آزمایشگاهی بر روی یک فرد، عبارت است از تشخیص افتراقی بیماری وی

با حداکثر دقت ممکن در معاینه بالینی. این کار معمولاً توسط پزشک و با جمع‌بندی شرایط فرد صورت می‌گیرد. به عنوان مثال برای تشخیص دیابت و یا سرطان (که در حیطه

* نویسنده مسئول. تلفن: ۰۲۳۱-۳۳۳۲۰۸۲، ۰۲۳۱-۳۳۳۱۵۵۱، نامبر: ۰۲۳۱-۳۳۳۱۵۵۱، E-mail: sadat-hashemi@sem-ums.ac.ir



شکل ۱. انواع مختلف منحنی‌های ROC

برای استفاده از منحنی ROC در ارزیابی یک تست تشخیصی، به‌کارگیری آماره A پیشنهاد شده است [۸]. این مساحت عبارت است از احتمال این که هرگاه متغیر تست به صورت تصادفی برای یک فرد بیمار و یک فرد سالم اندازه‌گیری شود، مقدار منتهی به ترتیب صحیح باشد (یعنی مثلاً برای فرد بیمار بزرگ‌تر از فرد سالم باشد و یا برعکس) [۸]. هرگاه تست قادر به تشخیص دقیق و بی‌نقص باشد، آنگاه مقادیر آن برای افراد سالم جمعیت پائین و برای افراد بیمار بالا خواهد بود (یا برعکس بسته به وضعیت بیماری). هرچه قدرت تشخیص تست بیش‌تر باشد، منحنی ROC بالای قطر مربع و به حالت ایده‌آل (مساحت یک) نزدیک‌تر خواهد بود. منحنی ROC تستی بدون توان تشخیص، منطبق بر قطر مربع و تستی با توان تشخیص معکوس در پایین این قطر قرار می‌گیرد [۴،۲]. شکل ۱ وضعیت‌های متفاوتی از منحنی ROC را نشان می‌دهد.

برای برآورد منحنی ROC روش‌های گوناگونی وجود دارد که براساس روش‌های پارامتری و ناپارامتری قرار دارند [۳،۶،۷،۸]. در نرم‌افزارهای آماری مشهور نظیر SAS، SPSS (نسخه ۹ به بعد) و STATA امکان برآورد منحنی ROC، محاسبه آماره A و انحراف استاندارد آن از طریق ناپارامتری وجود دارد. اما مشکل همه این نرم‌افزارها در ناتوانی آنها برای مقایسه دو منحنی ROC هم‌بسته است. این نوع ROCها هنگامی به‌وجود می‌آیند که بر روی هر فرد نمونه

پیش‌گیری سطح دوم قرار می‌گیرند، از ملاک‌ها و شاخص‌های مختلفی استفاده می‌شود. به طور کلی هرگاه C رده معین و مجزا در تشخیص داشته باشیم، آنگاه هر مورد نمونه با احتمال p_i ($\sum_{i=1}^c p_i = 1, i = 1, 2, \dots, c$) می‌تواند در گروه آم قرار بگیرد. ساده‌ترین حالت هنگامی است که هدف، تخصیص نمونه‌ها به یکی از دو گروه ممکن (سالم یا بیمار) باشد که در این صورت صفتی در مقیاس پیوسته و یا رتبه‌ای بر روی نمونه‌ای از افراد سالم و بیمار اندازه‌گیری می‌شود و سپس با انتخاب یک نقطه برش مناسب برحسب میزان حساسیت و ویژگی تست در آن نقطه و نیز مقدار تابع زیان اهمیت تشخیص، افراد سالم و بیمار مشخص می‌گردند؛ حال فرض کنید این شاخص‌ها و استانداردها برای بیماری خاصی موجود نباشند و محقق با بررسی‌های خود در پی آنها باشد. بنابراین اولین مسأله‌ای که با آن روبرو خواهد شد، عبارت است از ارزیابی هر یک از شاخص‌های تشخیصی و مقایسه آنها با یک‌دیگر به منظور دسترسی به بهترین آنها. چرا که توان تشخیص هر شاخص (تست) می‌تواند تحت تأثیر خطاهای گوناگون با دیگری متفاوت باشد. یک روش مؤثر و شناخته شده در جهت ارزیابی یک تست آزمایشگاهی (به طور کلی یک روش تشخیصی) که نتایج آن متغیری در مقیاس رتبه‌ای و یا کمی باشد، استفاده از منحنی ROC (Receiver operating characteristic) می‌باشد [۷،۴].

طبق نظریه، یک روش برای به دست آوردن منحنی ROC تجربی به این صورت می‌باشد. ابتدا نمونه‌ای از افراد سالم و بیمار انتخاب شده و تست مربوطه بر روی آنها اندازه‌گیری می‌شود، در ادامه دامنه تغییرات متغیر تست تعیین و سپس در نقاط برش متوالی در این دامنه حساسیت و مثبت کاذب (یک منهای ویژگی) در هر نقطه برش تعیین و نقاط به دست آمده به یک‌دیگر متصل می‌گردند. از این روش منحنی ROC به دست می‌آید که مساحت زیر آن (که ما برای سادگی، آن را آماره A نام نهاده‌ایم) بیان‌گر قدرت تشخیص تست می‌باشد.

$$(3) \quad \psi(x, y) = \begin{cases} 0 & y > x \\ 0.5 & y = x \\ 1 & y < x \end{cases}$$

آن گاه :

$$(4) \quad a = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n \psi(x_i, y_j)$$

برآوردی ناپارامتری از آماره A خواهد بود [۲]. لازم به ذکر است که تنها فرض آماری بر روی X ها و Y ها مربوط به مقیاس آنهاست و آنها بایستی حداقل در مقیاس رتبه‌ای باشند و توزیع آنها هیچ اهمیتی در استنباط‌های ما ندارد. هرگاه k تست وجود داشته باشد، آن‌گاه بردار $a = [a_1, a_2, \dots, a_k]$ برآوردی از A های منسوب به تست‌های مختلف خواهد بود ($A = [A_1, A_2, \dots, A_k]$). در این حالت ماتریس واریانس - کوواریانس این برآوردگرها عبارت است از [۱]:

$$(5) \quad S = \frac{1}{m} S_{10} + \frac{1}{n} S_{01}$$

که در آن Γ و S_{10} مؤلفه عبارت است از:

$$(6) \quad S_{10}^{rs} = \frac{1}{m-1} \sum_{i=1}^m [V_{10}^r(x_i) - a_r][V_{10}^s(x_i) - a_s]$$

و به طور مشابه Γ و S_{01} مؤلفه عبارت است از:

$$(7) \quad S_{01}^{rs} = \frac{1}{n-1} \sum_{j=1}^n [V_{01}^r(y_j) - a_r][V_{01}^s(y_j) - a_s]$$

که در آنها V ها بر حسب روش هوفدینگ محاسبه می‌شوند [۵].

برای آزمون فرضیه‌ای به صورت $H_0 : LA' = La'$ در برابر $H_1 : LA' \neq La'$ که در آن برداری از قیدها باشد. یک آزمون معنی‌داری را می‌توان به کمک آماره زیر انجام داد [۱]:

$$(8) \quad Z = \frac{La' - LA'}{\sqrt{LSL'}}$$

هم‌زمان چند تست مختلف اندازه‌گیری شده و بر مبنای هر تست یک منحنی ROC بنا نهاده شده باشد، که در این صورت یک هم‌بستگی طبیعی بین ROCها وجود خواهد داشت که صرف‌نظر کردن از آن می‌تواند منجر به نتایج گمراه‌کننده شود [۱]. در این مقاله دو هدف اصلی را دنبال کرده‌ایم:

(۱) ارائه روش مناسب برای مقایسه منحنی‌های ROC

هم‌بسته به منظور تعیین برترین تست‌ها

(۲) ارائه نرم‌افزار مناسبی که تحلیل مذکور را انجام دهد.

مواد و روش‌ها

فرض کنید نمونه‌ای متشکل از N فرد داشته باشیم که m تای آنها بیمار و $n=N-m$ تای آنها سالم باشند. هم‌چنین فرض کنید X_i ($i=1,2,\dots,m$) مقادیر متغیر تشخیص برای بیماران و Y_j ($j=1,2,\dots,n$) مقادیر متغیر تشخیص برای افراد سالم باشد. به ازای هر عدد حقیقی Γ حساسیت تجربی تست برای نقطه برش Γ عبارت است از:

$$(1) \quad Sen(r) = \frac{1}{m} \sum_{i=1}^m I(X_i \geq r)$$

و ویژگی تجربی در همین نقطه عبارت است از:

$$(2) \quad Spe(r) = \frac{1}{n} \sum_{j=1}^n I(Y_j < r)$$

که در آنها $I(A)=1$ هرگاه شرط A برقرار باشد و در غیر این صورت $I(A)=0$.

با تغییر Γ بر روی فضای ممکن، مقادیر تشخیص منحنی ROC تجربی از رسم $Sen(r)$ بر حسب $1-Spe(r)$ به دست می‌آید. هرگاه Γ بزرگ‌تر از حداکثر مقدار ممکن متغیر تست باشد، منحنی از (۰ و ۰) می‌گذرد و به‌طور یک‌نواخت با کاهش Γ به سمت نقطه (۱ و ۱) افزایش می‌یابد. هرگاه تابع Ψ به صورت زیر باشد:

بحث و نتیجه گیری

در قسمت نتایج ابتدا چنین به نظر می آید که تست t_1 به دلیل داشتن مقدار A ی بزرگ تر نسبت به t_2 برتری داشته باشد، اما در عمل نتیجه آزمون بر مبنای روش رایج شده در این جا چنین تفاوتی را نشان نمی دهد. همان طور که از نتایج مشهود است، صرف نظر کردن از همبستگی بین A ها می تواند منجر به نتایج گمراه کننده شود (البته نه لزوماً). زیرا در برخی مواقع اختلاف آماره ها آن قدر زیاد (کم) است که اثر کواریانس را در آزمون فرضیه از بین می برد و چه کواریانس در نظر گرفته شود و یا در نظر گرفته نشود، باز هم نتیجه آزمون ها یکسان و منجر به رد (پذیرش) فرض صفر خواهد شد. با این حال، مشکل هنگامی پیش می آید که صرف نظر کردن از همبستگی ها گمراه کننده باشد و منجر به استنباط غلط شود (نظیر مدل شبیه سازی شده)، در این صورت باید حتماً همبستگی در نظر گرفته شود، زیرا در نتیجه آزمون نقش تعیین کننده خواهد داشت. این نظیر حالتی است که ما هنگامی که داده ها واجد شرایط استفاده از آزمون t ی زوجی باشند، از آزمون t ی مستقل برای مقایسه استفاده کرده باشیم. این مطلب بدون هیچ نیازی به شبیه سازی و تنها با دقت در رابطه ۹ مشهود می باشد.

برای دخیل کردن اثر همبستگی بین ROCها کارهای دیگری نیز انجام شده که مبتنی بر روش های رگرسیونی لجستیک دوحالته و رتبه ای هستند و در آن ها مقادیر مختلف نقاط برش در دامنه مقادیر متغیرهای تست به عنوان متغیرهای پیش بین گر بیماری در نظر گرفته شده و از اثر متقابل آن ها در بیان همبستگی استفاده می شود [۹، ۱۱]، برآورد منحنی های ROC در کنار داده های گم شده نیز مورد بررسی قرار گرفته است (اما بدون در نظر گرفتن همبستگی های بین آن ها) [۱۰]؛ اما روش رایج شده در اینجا با تلفیقی از روش رایج شده در (۱۰) قابل استفاده برای داده های گم شده نیز می باشد.

از آن جا که برای ارزیابی و مقایسه تست های تشخیصی هیچ روش آماری مناسبی به جز استفاده از منحنی های ROC وجود ندارد، لذا روشی که در این مقاله ارائه شده با توجه به

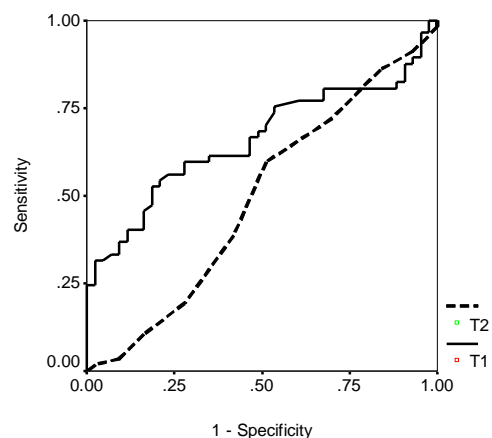
که دارای توزیع نرمال استاندارد است [۱]. در حالتی که هدف مقایسه A_1^1 و A_1^2 مربوط به دو تست باشد آماره ۸ به شکل ساده زیر در می آید:

$$(9) \quad Z = \frac{a_1 - a_2}{\sqrt{S_{11} + S_{22} - 2S_{12}}}$$

که در آن S_{ij} عناصر ماتریس $S_{2 \times 2}$ هستند. برای مقایسه عملی با استفاده از زبان Delphi5، نرم افزار مناسبی به نام ROC را تهیه کرده ایم که در آدرس www.sem-ums.ac.ir موجود است.

نتایج

برای پیش برد اهدافمان از داده های شبیه سازی شده (که در همان آدرس اینترنتی ذکر شده با عنوان ROC.dat وجود دارد) استفاده کردیم این داده ها شامل نمونه ای متشکل از ۴۳ فرد سالم و ۵۷ فرد بیمار به همراه مقادیر دو تست t_1 و t_2 برای آن ها می باشد، شکل ۲ منحنی های ROC مربوطه را نشان می دهد. مقادیر محاسبه شده برای این ROCها عبارتند از: $A_1 = 0.66$ ، $SE(A_1) = 0.054$ ، $A_2 = 0.491$ ، $SE(A_2) = 0.060$ در این حالت با صرف نظر کردن از همبستگی بین A_1 و A_2 داریم $Z = 2.1$ که در سطح معنی داری ۵٪ منجر به رد تساوی A_1 و A_2 خواهد شد. حال با در نظر گرفتن همبستگی A_1 و A_2 ، $Z = 1.92$ به دست می آید که در همان سطح معنی داری منجر به پذیرش تساوی A_1 و A_2 می شود.



شکل ۲. منحنی های ROC مربوط به مثال شبیه سازی شده

confidence intervals rating method data. *J Math Psych*, 1969; 6: 487.

[4] Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 1982 Apr; 143(1):29-36.

[5] Hoeffding W. A class of statistics with asymptotically normal distribution. *Ann Math Statist*, 1948; 19: 293-325.

[6] Metz CE. Basic principles of ROC analysis. *Semin Nucl Med*, 1978 Oct; 8(4):283-98.

[7] Metz CE, Wang P-L, Kronman HB. A new approach for testing the significance of differences between ROC curves measured from correlated data. In: Deconinck F. (editor). *Information processing in medical imaging*. The Hague: Nijhoff, 1984, p. 432-445.

[8] Swets JA, Pickett RM. *Evaluation of diagnostic systems: methods from signal detection theory*. New York: Academic Press, 1982.

[9] Toledano AY, Gatsonis C. Ordinal regression methodology for ROC curves derived from correlated data. *Stat Med*, 1996 Aug 30; 15(16):1807-26.

[10] Vida S. A computer program for non-parametric receiver operating characteristic analysis. *Comput Methods Programs Biomed*, 1993 Jun; 40(2):95-101.

[11] Zhou XH, Gatsonis CA. A simple method for comparing correlated ROC curves using incomplete data. *Stat Med*, 1996 Aug 15; 15(15):1687-93.

سادگی و دور بودن از محاسبات پیچیده آماری می‌تواند به سادگی توسط محققین به عنوان یک سیستم پشتیبان در تصمیم‌گیری به کار گرفته شده و در مواردی که هدف، تعیین بهترین شاخص‌های تشخیصی باشد، مورد استفاده قرار گیرد.

منابع

[۱] سادات‌هاشمی مهدی. معماری شبکه‌های عصبی مصنوعی برای مدل‌بندی

پاسخ‌های چندمتغیره دوحالته و به‌کارگیری آن‌ها در پیش‌بینی حاملگی‌های

ناخواسته. پایان‌نامه دکترای تخصصی در رشته آمار زیستی، تهران: دانشگاه تربیت

مدرس، ۱۳۸۲: ۱۳۰-۱۲۶.

[2] Bamber D, The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *J Math Psychol*, 1975; 12: 387-415.

[3] Dorfman DD, Alf E. Maximum likelihood estimation of parameters of signal detection theory and determination of

