

روش کار برنامه‌ی "نمودارهای توان اجرا"

Power Function Graphs

توجه:

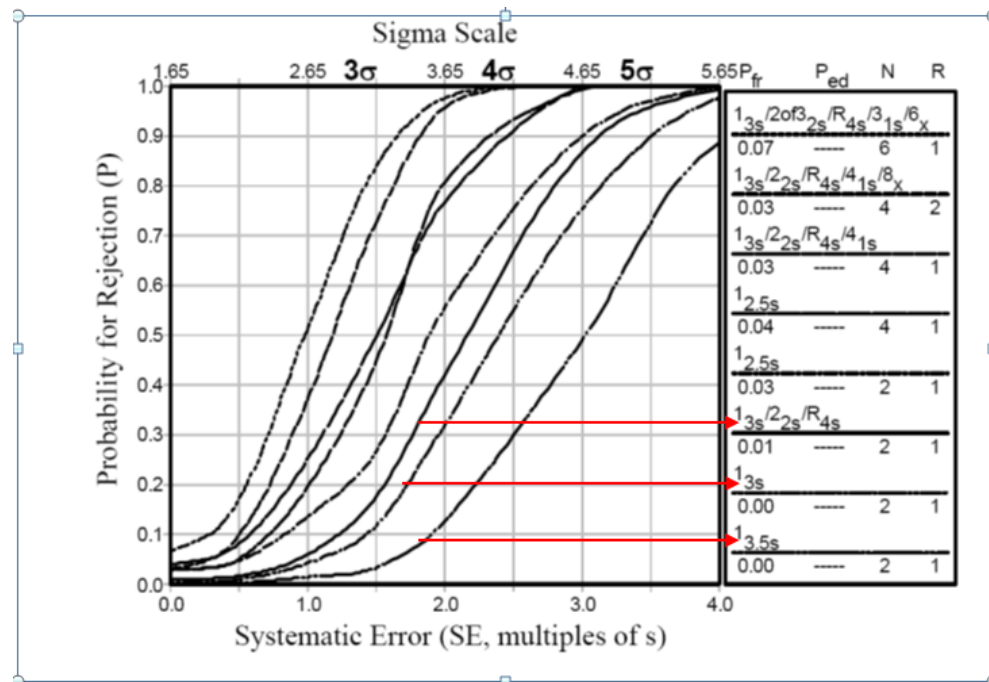
(۱) اجرای این حسابگر به ویرایش ۲۰۱۰ یا بالاتر برنامه‌ی Excel نیاز دارد.

(۲) چون این حسابگر در برنامه‌ی Excel تهیه شده است، ممکن است ابعاد صفحه‌ی باز شده همسان با صفحه‌ی نمایشگر نباشد. در این صورت لازم است پس از باز کردن برنامه با کاستن یا افزودن بر اندازه‌ی صفحه آن را طوری تنظیم کرد که مناسب نمایشگر باشد؛ صفحه را طوری تنظیم کنید که در ستون n دست کم تا ردیف ۲۶ دیده شود. برای این که هر بار مجبور به این کار نباشیم، می‌توان پیش از باز کردن برنامه، تیک Read-only را برداشت، صفحه را باز و اندازه‌ی آن را تنظیم و ذخیره کرد و پس از بستن صفحه دوباره Read-only را تیک زد.

❖ توضیح کوتاهی در باره‌ی نمودارهای توان اجرا

برای پایش کیفیت روش‌های سنجش، به سامانه‌هایی برای خطایابی نیاز داریم. معیارهای "پایش کیفیت آماری" (Statistical Quality Control; SQC) همان سامانه‌های خطایابی ما هستند. توانایی یک معیار SQC در شناسایی خطا، تابعی است از اندازه‌ی خطا؛ به این معنا که هرچه اندازه‌ی خطا بزرگتر باشد، احتمال شناسایی آن به وسیله‌ی آن معیار خاص بیشتر می‌شود.

یک "نمودار توان اجرا" برای یک معیار SQC خاص، نموداری است که رابطه‌ی بین "اندازه‌ی خطا" و "احتمال خطایابی" آن معیار را نمایش می‌دهد. در زیر مجموعه‌ای از این نمودارها برای شناسایی خطای سامانمند (Systematic Error) نمایش داده شده است (مجموعه‌ی زیر مانند بسیاری دیگر از این نمودارها حاصل زحمات پروفیسور جیمز ا. وستگارد است):



در این شکل ۸ خط منحنی دیده می‌شود که از چپ به راست هر یک از آن‌ها مربوط است به یکی از معیارهای درون مستطیل سمت راست به ترتیب از پایین به بالا. محور عرض‌های این نمودار، احتمال شناسایی

خطا (Probability of Error Detection; Ped) را نشان می‌دهد و محور طول‌ها برابر خطای سامانمند مورد نظر با واحد "انحراف معیار" است.

برای این که بدانیم کدام معیار به کار پایش کیفیت یک روش می‌آید ابتدا باید بررسی کنیم که با بروز چه مقدار خطا، باید زنگ هشدار ما باید به صدا درآید. این مقدار از خطا که "خطای سامانمند بحرانی؛ Critical SE" نامیده می‌شود بستگی دارد به "ویژگی‌های اجرایی" روش ما [شامل انحراف معیار (CV) و نامیزانی (B)]، و "خطای کل مجاز (ATE)" برای سنجش آن آنالیت. با دانستن این مشخصات می‌توان از رابطه‌ی زیر خطای سامانمند بحرانی را حساب کرد:

$$SE_{crit} = [(ATE - B)/CV] - 1.64$$

مثلا اگر برای سنجش گلوکز از روشی با انحراف معیار برابر ۱.۵% و نامیزانی ۲.۵% استفاده کنیم، با در نظر گرفتن این که خطای کل مجاز برای سنجش گلوکز در جدول CLIA برابر ۱۰% است، خطای سامانمند بحرانی ما برابر ۳.۳۵ انحراف معیار خواهد بود:

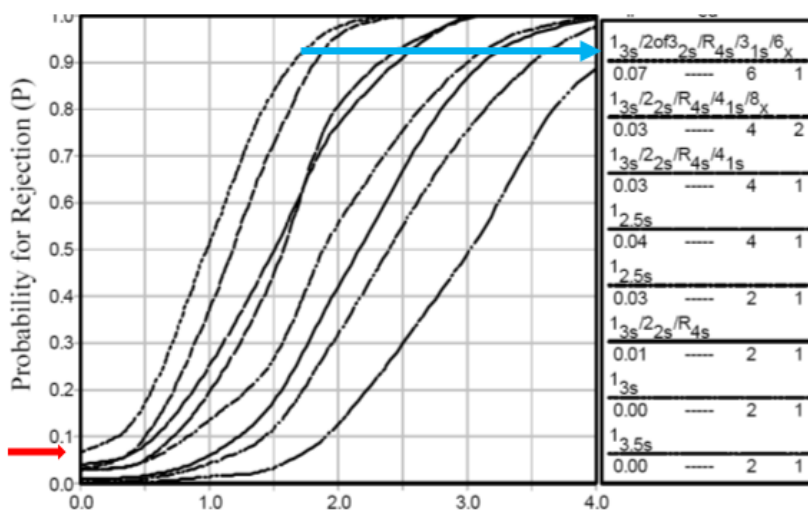
$$SE_{crit} = [(10 - 2.5)/1.5] - 1.64 = 3.4 \text{ SD}$$

در سطح تصمیم ۱۰۰ میلی‌گرم بر دسی‌لیتر، انحراف معیار روش ما برابر ۱.۵ میلی‌گرم بر دسی‌لیتر است:

$$SD = 100 * CV = 100 * 1.5\% = 1.5 \text{ mg/dL}$$

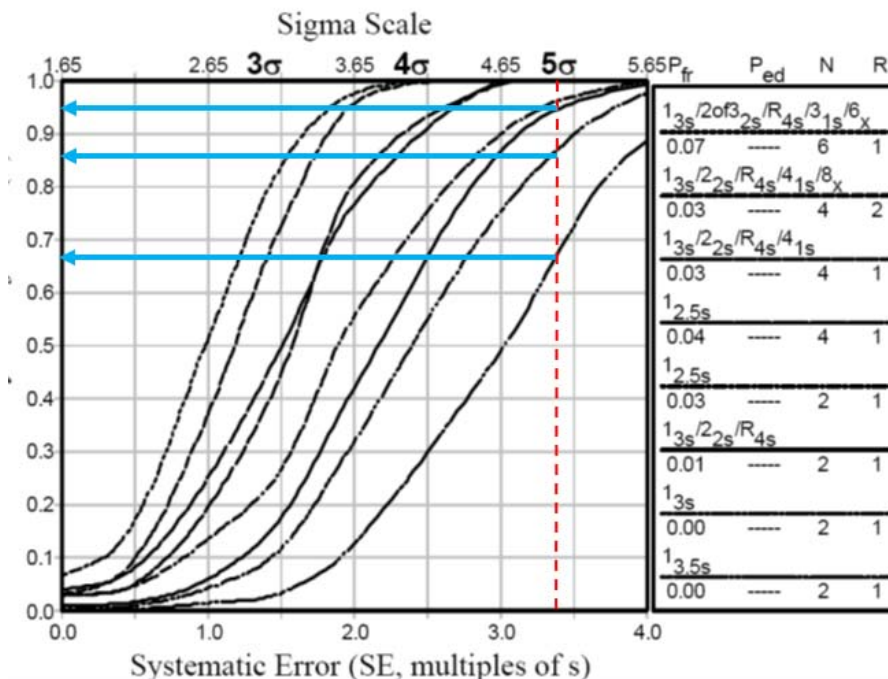
بنا بر این ما به معیاری برای پایش کیفیت نیاز داریم که وقتی کالیبر دستگاه ما در سطح ۱۰۰ به اندازه‌ی ۵ میلی‌گرم بر دسی‌لیتر ($3.35SD = 3.4 * 1.5 = 5.1 \text{ mg/dL}$) جابجا شد ما را خبر کند، و البته پیش از آن ما را خبر نکند چون هنوز خطای مهمی به شمار نمی‌آید.

حال باید به نمودار بالا مراجعه کنیم و قانون مناسب را انتخاب کنیم. اما پیش از آن باید از محدودیتی که در عمل داریم آگاه شویم: اگرچه بسیار خوب است که معیاری را انتخاب کنیم که هنگام بروز خطایی به بزرگی خطای بحرانی، با احتمال ۱۰۰% بتواند ما را خبر کند اما همواره نمی‌توان چنین معیاری را برگزید؛ زیرا گاه چنین حساسیت بالایی در شناسایی خطا سبب خواهد شد که در موارد زیادی بدون وجود خطا، اعلان خطا دریافت کنیم که در واقع "رد کاذب؛ False Rejection" است. احتمال رد کاذب (Probability of False Rejection; Pfr) برای یک معیار پایش کیفیت برابر است با جایی که نمودار توان آن با محور عرض‌ها برخورد می‌کند؛ یعنی جایی که خطای سامانمند ما صفر است. مثلا با مراجعه به شکل زیر دیده می‌شود که برای اولین نمودار از طرف راست که مربوط است به معیار چندقانونی $1_{3s}/2_{of}3_{2s}/R_{4s}/3_{1s}/6_x$ با ۶ کنترل در هر دور (۲ سنجش از هر سه سطح) برابر ۰.۰۷ یا ۷% است (پیکان قرمز):



حال آنکه برای دو معیار آخر مستطیل، احتمال رد کاذب برابر صفر است. در مستطیل سمت راست، احتمال رد کاذب هر معیار زیر عنوان Pfr نمایش داده شده است. در عمل باید معیاری را انتخاب کنیم که دست کم ۹۰٪ توان خطیابی داشته باشد ($Ped \geq 90\%$) و همزمان احتمال رد کاذب آن کمتر از ۵٪ باشد ($Pfr \leq 5\%$).

حال با آگاهی از این محدودیت برگردیم سر مثالمان؛ بناست برای SEcrit برابر $3.35 SD$ معیار پایش کیفیت پیدا کنیم. برای این کار روی محور طول‌ها عدد ۳.۳۵ را پیدا می‌کنیم و از آن نقطه یک خط عمودی رسم می‌کنیم (خطچین قرمز در شکل زیر):



چنانچه از محل برخورد خطچین قرمز با هر یک از نمودارها خطی افقی به سمت محور عرض‌ها رسم کنیم، محل برخورد آن خط با محور عرض‌ها Ped آن معیار را نمایش خواهد داد (پیکان‌های آبی). برای مثال، پایین‌ترین معیار در مستطیل حدود ۶۸٪ توان خطیابی دارد که کمتر از ۹۰٪ است. معیار دوم از پایین حدود ۸۷٪ توان خطیابی دارد که باز هم اندکی کم است. معیارهای سوم و چهارم حدود ۹۵٪ توان خطیابی دارند که خوب است و معیارهای بالاتر ۱۰۰٪ توان خطیابی دارند. برای مثال بالا، باید یکی از معیارهای سوم یا چهارم از پایین را انتخاب کنیم؛ سراغ معیارهای بالاتر نمی‌رویم زیرا اگرچه توان خطیابی بیشتری دارند اما احتمال رد کاذب آن‌ها نیز بالاست و همچنین سخت‌تر و پرهزینه‌تر هستند.

مستطیل سمت راست اطلاعات دیگری نیز در بردارد:

- اعداد نوشته شده زیر ستون N نشان‌دهنده تعداد کنترل در هر دور است. مثلا برای اولین معیار از بالا باید در هر دور ۶ کنترل بگذاریم (۲ سنجش از هر سه سطح) در حالی که برای دو معیار بعدی ۴ کنترل در هر دور و برای ۴ معیار پایینی تنها ۲ کنترل در هر دور کافی است.
- اعداد نوشته شده زیر ستون R بیانگر این است که برای تفسیر آن معیار، نتایج کنترل‌ها در چند دور باید با هم مقایسه شوند. در مورد این نمودارها R برای همه‌ی معیارها برابر ۱ است و این به این معناست که در هر دور تکلیف پذیرش یا رد آن دور با کنترل‌های گذاشته شده در همان دور تعیین می‌شود و کاری به دورهای پیشین نداریم. چنانچه R برابر ۲ باشد باید نتایج کنترل دو دور پی‌درپی را در نظر گرفت.

برای مثال بالا یا باید از معیار سوم از پایین استفاده کنیم، $1_{3s}/2_{2s}/R_{4s}$ ، که اگرچه نسبت به معیار چهارم احتمال رد کاذب کمتری دارد ($Pfr = 1\%$) اما تفسیر آن به خاطر چدقانونی بودن کمی سخت است. قانون چهارم، $1_{2.5s}$ ،

ساده‌تر است اما Pfr آن ۳% است. انتخاب بین یکی از این دو قانون بستگی دارد به مهارت کاربر در تفسیر معیارها. البته اگر در آینده دیدیم که پایداری دستگاه ما برای آن روش خوب است و احتمال به خطا افتادن آن کمتر از ۱% است می‌توانیم به ۸۷% توان خطیابی بسنده کنیم و از معیار دوم از پایین، معیار $1_{2.5s}$ با Pfr برابر صفر استفاده کنیم.

آنچه گفته شد مربوط بود به خطای سامانمند. برای شناسایی خطای اتفاقی باید به نمودارهایی که ویژه‌ی خطای اتفاقی رسم شده‌اند مراجعه کرد. اما در حال حاضر با توجه به استفاده از دستگاه‌های خودکار، خطایی که باید به آن توجه شود خطای سامانمند است. CLSI نیز در دستور کار 24-A3 C برای پایش کیفیت، مجموعه‌ای از معیارهای SQC برای شناسایی خطای سامانمند را انتخاب و در قالب یک نمودار توان ارائه کرده است.

❖ روش کار برنامه

این برنامه بر اساس نمودار توان پیشنهادی CLSI طراحی شده است و بر اساس اطلاعات وارد شده محاسبات لازم را انجام و معیار پایش کیفیت مناسب را پیشنهاد می‌کند.

برای استفاده از این برنامه به روش زیر کار شود:

۱) در جدول سمت چپ، از بالا به پایین به ترتیب خطای کل مجاز، انحراف معیار و نامیزانی را وارد کنید. مثلاً برای یک روش سنجش PTT با $B = 2\%$ و $CV = 3\%$ با در نظر گرفتن خطای کل مجاز برابر ۱۵% (بر اساس جدول CLIA) اطلاعات را به شکل زیر وارد می‌شود:

DATA	
ATE	15
Bias	2
CV	3
SM	4.333333
%AD	
Min. Ped	
$\Delta SE (SD)$	
Assured Q	

با وارد کردن این داده‌ها عیار سیگمای این روش برابر ۴.۳ حساب شده نمایش داده می‌شود.

(۲) در مقابل عنوان %AD (Allowable Defect %) معین کنید که اجازه داریم چند درصد جواب‌های مان بیرون از محدوده‌ی خطای کل مجاز باشد. روند معمول کنونی این است که اجازه داریم ۵% جواب نادرست تولید کنیم. این ۵% بر هیچ دلیل خاصی استوار نیست و بیشتر یک سنت است؛ آنطور که پروفیسور وستگار می‌گوید باید تلاش کنیم به سمت ۱% و سپس ۰.۱% حرکت کنیم. برای این مثال همان ۵% معمول را در نظر می‌گیریم.

(۳) در مقابل عنوان Min. Ped تعیین کنید که انتظار دارید توان خطایابی معیار پایش کیفیت انتخاب شده چقدر باشد. همانطور که در بالا گفتیم خطایابی باید دست‌کم ۹۰% باشد.

(۴) پس از وارد کردن %AD و Min. Ped، جدول سمت چپ کامل می‌شود:

DATA	
ATE	15
Bias	2
CV	3
SM	4.333333
%AD	5
Min. Ped	90
Δ SE (SD)	2.7
Assured Q	1.64

در مقابل عنوان Δ SE (SD) خطای سامانمند بحرانی برابر ۲.۷ نمایش داده می‌شود که به این معناست که ما نیاز داریم هنگامی که کالیبر روش سنجش ما به اندازه‌ی ۲.۷ SD جابجا شد خبر شویم. در ردیف آخر در مقابل عنوان Assured Q عدد ۱.۶۴ دیده می‌شود که به این معناست که ما کیفیت ۱.۶۴ سیگما را تضمین می‌کنیم.

(۵) در جدول میانی معیارهای گوناگون پایش کیفیت دیده می‌شود و برنامه بررسی می‌کند که آیا با توجه به داده‌های وارد شده در جدول سمت چپ می‌تواند از میان آن‌ها برای ما معیار مناسب انتخاب کند یا نه. معیار پیشنهادی با حرف R قرمز رنگ مشخص می‌شود. در مورد مثال بالا وضعیت زیر نمایش داده خواهد شد:

CRITERIA	N	R	P _{fr}	P _{ed}	Line
1 _{3s} /2of3 _{2s} /R _{4s} /3 _{1s} /6 _x	6	1	7%	99.1%	
1 _{3s} /2 _{2s} /R _{4s} /4 _{1s}	4	1	5%	92.2%	
1 _{2.5s}	4	1	4%	93.8% R	
1 _{2.5s}	2	1	4%	79.1%	
1 _{3s} /2 _{2s} /R _{4s}	2	1	2%	72.7%	
1 _{3s}	2	1	2%	61.9%	
1 _{3.5s}	2	1	1%	39.3%	
1 _{3s}	1	1	0%	37.0%	

R: Recommended Criteria

در این جدول برای مثال ما معیار سوم از بالا پیشنهاد می‌شود؛ یعنی معیار 1_{2.5s} با گذاشتن ۴ کنترل (۲ سنجش از هر سطح) در هر دور. با چنین معیاری حدود ۹۴٪ توان خطایابی خواهیم داشت و البته حدود ۴٪ هم احتمال رد کاذب وجود دارد. عدد ۱ در ستون R هم به این معناست که تکلیف پذیرش یا رد یک دور با بررسی نتیجه‌ی کنترل‌ها در همان دور معلوم می‌شود و کاری به دورهای پیشین نخواهیم داشت. نوارهای رنگی جلوی هر ردیف راهنمای این است که در نمودار توان اجرا که در سمت راست نمایش داده می‌شود، منجنی مربوط به آن معیار کدام است.

(۶) نمودارهای توان نمایش داده شد در سمت راست برای آن است که بتوان آنچه را که برنامه حساب کرده است به صورت چشمی نیز بررسی کرد. خطچین قرمز در این نمودار از محل خطای سامانمند بحرانی رسم می‌شود.

اگر مثال PTT را برای روشی با مشخصات $B = 1\%$ و $CV = 2\%$ بررسی کنیم معیار پیشنهادی پایینترین ردیف جدول خواهد بود؛ یعنی معیار 1_{3s} با گذاشتن تنها ۱ کنترل در هر دور، با توان خطایابی ۱۰۰٪ و با احتمال رد کاذب صفر در صد!

علت این که برای روش قبلی باید ۴ کنترل در هر دور بگذاریم و علاوه بر آن ۴٪ موارد هم اعلان رد کاذب دریافت کنیم، در حالی که با این روش معیاری بسیار آسانتر، مطمئنتر و کم هزینهتر پیشنهاد می شود، در اختلاف کیفیت روش هاست. عیار کیفیت روش اول ۴.۳ سیگماست در حالی که عیار کیفیت روش دوم ۷ سیگماست. در نظر داشته باشید که با هر دو روش می خواهیم کیفیت ۱.۶۴ سیگما را تضمین کنیم (یعنی تضمین کنیم که حداکثر ۵٪ از جواب های PTT بیماران بیشتر از ۱۵٪ با مقدار واقعی شان اختلاف خواهند داشت). این به آن می ماند که بخواهیم قبولی دو دانش آموز در آخر سال را، در حالی که اولی با نمره ۱۳ در ثلث اول قبول شده است و دومی با نمره ۱۹، تضمین کنیم. با در نظر گرفتن این که نمره قبولی آخر سال ۱۰ است، مطمئناً برای تضمین قبولی دانش آموز اول باید "مراقبت" سخت گیرانه ای به کار بندیم چرا که فاصله ۱۳ تا ۱۰ بسیار کم است و حاشیه ای اطمینان کمی داریم. اما در مورد دانش آموز دوم معیار پایش خیلی سختی نیاز نداریم چرا که حاشیه ای امنیت بین ۱۹ تا ۱۰ خیلی بزرگ است. در مورد روش ها هم چنین است. حاشیه ای امنیت بین ۷ تا ۱.۶۴ بسیار بزرگتر است از حاشیه ای امنیت بین ۴.۳ تا ۱.۶۴. بنا بر این؛ روشی با کیفیت ۷ سیگما پایش کمتری نیاز دارد.

برای روش ۷ سیگما حتی می توانیم نمره قبولی بالاتری را تضمین کنیم. اگر مقابل AD% به جای ۵ بگذاریم ۱ (یعنی بگوییم می خواهیم تضمین کنیم که حداکثر ۱٪ از جواب های PTT بیماران بیش از ۱۵٪ با مقدار واقعی اختلاف خواهد داشت؛ نه ۵٪)، خواهیم دید که باز هم همان معیار آسان 1_{3s} پیشنهاد می شود. مقابل عنوان Assured Q عدد ۲.۳۳ به نمایش در می آید یعنی ما داریم کیفیت ۲.۳۳ سیگما را تضمین می کنیم. حتی اگر مقابل AD% عدد ۰.۱ را که پروفسور وستگارد پیشنهاد می کند بگذاریم (یعنی بگوییم می خواهیم تضمین کنیم که حداکثر ۰.۱٪ از جواب های PTT بیماران بیش از ۱۵٪ با مقدار واقعی اختلاف خواهد داشت؛ و این یعنی از هر هزار جواب PTT که گزارش می کنیم حد اکثر یکی برای کاربرد بالینی نادرست خواهد بود)، معیار سوم جدول از پایین پیشنهاد می شود: معیار 1_{3s} با دو کنترل در هر دور، خطایابی تقریباً ۹۶٪ و رد کاذب تنها ۲٪. یعنی با کمی هزینه ای بیشتر می توانیم به جای کیفیت ۱.۶۴ سیگما، کیفیت ۳.۰۹ سیگما را تضمین کنیم. این یعنی این که با کمی مراقبت بیشتر می توانیم تضمین کنیم که دانش آموز شماره ۲ در بالا (نمره ۱۹ در ثلث اول) در آخر سال با نمره ای مثلاً ۱۵ قبول خواهد شد.

روش اول ($B = 2\%$ و $CV = 3\%$) را برای AD% های ۱ و ۰.۱ بررسی کنید و نتیجه را ببینید؛ ببینید که اگر بخواهیم تضمین کنیم در آخر سال دانش آموز یکم (نمره ۱۳ در ثلث اول) مثلاً با نمره ۱۲.۵ قبول خواهد شد (مثلاً معادل $AD = 1$) چه مشقتی خواهیم داشت! و نیز ببینید که اگر در جستجوی معیار مراقبتی باشیم برای تضمین این که این دانش آموز در آخر سال با نمره ۱۵ قبول خواهد شد (مثلاً معادل $AD = 0.1$) چه جوابی از برنامه خواهیم گرفت.

پیروز باشید

حسن بیات، ۱۳۹۲/۹/۲۱