

معرفی روشها و آزمون های آماری مختلف، در بررسی و انتخاب ویژگی ها

نگارش: گلناز بغدادی

فهرست مطالب

۱	مقدمه	۱
۱	مشاهده توزیع گروه ها	۱
۲	۱,۲ هیستوگرام	۲
۳	۲,۲ نمودار مستطیلی	۳
۶	۳,۲ نمودار پراکندگی	۶
۷	۳ آزمون های معنی دار بودن	۷
۸	۱,۳ فرض صفر و مخالف صفر	۸
۸	۲,۳ انواع خطاها در یک آزمون فرض	۸
۹	۳,۳ سطح معنی دار بودن	۹
۹	۴,۳ ارزیابی فرض نرمال بودن	۹
۱۰	۵,۳ تبدیلاتی برای نرمال کردن تقریبی	۱۰
۱۰	۶,۳ آنالیز واریانس یکطرفه	۱۰
۱۴	۷,۳ روشهای مقایسه چندگانه	۱۴
۱۵	۴ آنالیز واریانس چند متغیره یکطرفه	۱۵
۱۶	۱,۴ آزمون t با نمونه های جفت	۱۶
۱۸	۵ منحنی ROC و پارامتر AUC	۱۸
۲۲	۶ تبدیل Z-Score	۲۲
۲۵	۷ مراجع	۲۵

۱. مقدمه

پیدا کردن زیر مجموعه ای از ویژگی ها، از یک مجموعه بزرگ، مسئله ای است که در بسیاری از زمینه های مطالعاتی پیش می آید. از آنجایی که افزایش تعداد ویژگیها هزینه محاسباتی یک سیستم را افزایش میدهد، طراحی و پیاده سازی سیستمها با کمترین تعداد ویژگی ضروری به نظر می رسد. از طرف دیگر توجه به این موضوع بسیار مهم است که، باید زیر مجموعه موثری از ویژگیها انتخاب شود که کارایی قابل قبولی برای سیستم ایجاد کند. این موضوع ما را به سمتی هدایت می کند که از روشهای جستجو، برای پیدا کردن زیر مجموعه ای بهینه از ویژگی ها، استفاده کنیم. همچنین انتخاب ویژگی های مناسب می تواند تاثیر قابل توجهی بر نرخ بازشناسی درست الگوریتم های طبقه بندی نیز داشته باشد. در ابتدای طبقه بندی مشخص نیست که کدام زیرمجموعه از ویژگی ها بیشترین تمایز را برای کلاسهای مورد مطالعه ایجاد می کنند و از طرف دیگر امکان بررسی تمام زیر مجموعه های موجود، از نظر صرف وقت، امکان پذیر و مقرون به صرفه نیست. هدف اصلی از انتخاب ویژگی، کاهش بعد بردار ویژگی در طبقه بندی است بطوریکه نرخ طبقه بندی قابل قبولی نیز حاصل شود. در این شرایط، ویژگی هایی که قدرت تمایز کمتری دارند، حذف شده و یکسری از ویژگیها که شامل اطلاعات مناسبی برای تمایز کلاسهها هستند، باقی می ماندند [۱]. تاکنون روشهای زیادی برای استخراج و انتخاب ویژگی ارائه شده است. بسیاری از روشهای متداول در مرجع [۲] مرور و بررسی شده اند. در این مطالعه به معرفی و بررسی آزمونها و روشهای آماری مختلف، در انتخاب ویژگی های بهینه پرداخته شده است.

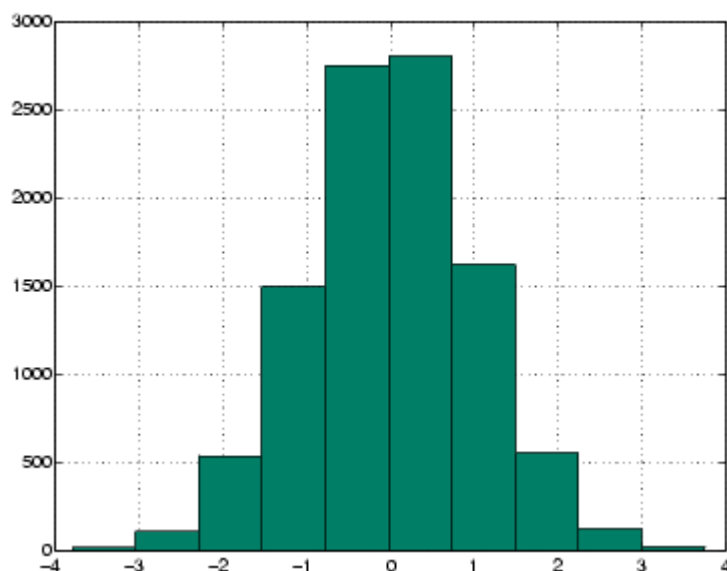
در طبقه بندی گروه های مختلف به کمک یک یا چند ویژگی، معمولاً جداسازی انجام نمی شود مگر اینکه بردارهای میانگین جامعه ها به طور معنی داری باهم اختلاف داشته باشند. اگرچه تفاوتی معنی دار ظاهری به طور خودکار کارایی ویژگی های مختلف را نتیجه نمی دهد، ولی آزمون کردن اولین قدم است. اگر هیچ تفاوت معنی داری دیده نشود، احتمالاً ساختن طبقه بندی کننده اتلاف وقت است [۳]. برای اجرای این آزمون ها، در علم آمار روشهای متعددی برای مقایسه بردارهای میانگین جوامع یک یا چند متغیره ارائه شده است که در ادامه به معرفی آنها پرداخته خواهد شد.

۲. مشاهده توزیع گروه ها

در علم آمار روشهای متعددی وجود دارد که به بررسی و توصیف متغیر ها، به صورت دیداری، کمک می کنند. هر کدام از این روشهای دیداری به نوعی توزیع دادگان را نمودار می کنند. در ادامه تعدادی از آنها معرفی خواهد شد.

۱.۲ هیستوگرام

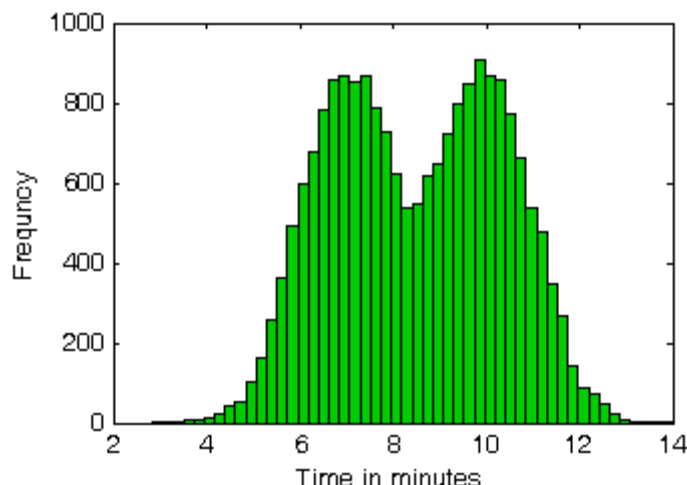
هیستوگرام کمک می کند تا نحوه تجمع و توزیع مقادیر داده های مربوط به نمونه ها مشخص شود. زمانی که تعداد داده ها بیشتر است، هیستوگرام اطلاعات موجود در داده ها را بهتر نمایش می دهد. محور افقی نمودار هیستوگرام بیانگر محدوده مقادیر داده ها است و محور عمودی این نمودار، نشان می دهد که به ازای هر مقدار از داده ها، چه تعداد نمونه وجود دارد. به عنوان مثال هیستوگرام رسم شده در شکل زیر را در نظر بگیرید:



شکل ۱: نمودار هیستوگرام

با توجه به این شکل می توان گفت اکثر داده ها در محدوده بین -۱ تا ۱ هستند. به این ترتیب این نمودار می تواند نشان دهنده موارد زیر باشد [۴]:

- نشان می دهد که آیا تجمع نمونه ها به سمت یک مقدار خاص است یا خیر؟
- مشخص میکند که آیا توزیع داده ها متقارن است یا خیر؟
- وجود دسته های جداگانه در مقادیر نمونه ها را نشان می دهد. برای مثال زمانیکه در نمودار هیستوگرام داده ها، دو پیک (یا بیشتر) وجود داشته باشد، ممکن است بیانگر وجود دو گروه متفاوت باشد. به عنوان نمونه در شکل زیر نمودار داده های مربوط به زمان رسیدن به خط پایان در یک مسابقه دو، از ۲۰۰۰ نفر رسم شده است.



شکل ۲: هیستوگرام داده های مربوط به زمان رسیدن به خط پایان در یک مسابقه دو، از ۲۰۰۰۰ نفر

با توجه به اینکه در توزیع دادگان دو پیک مشاهده می شود، بیانگر وجود دو گروه متفاوت در دادگان است. ممکن است به نظر برسد که این دو گروه متفاوت مربوط به مردها و زن ها است، در این موارد ممکن است بخواهید که داده ها را برای مردها و زن ها به طور جداگانه بررسی نمایید.

* در محیط مطلب برای رسم این نمودار از دستور *hist* استفاده می شود.

۲.۲. نمودار مستطیلی^۱

نمودارهای ستونی برای نمایش اطلاعات خلاصه ای از گروه ها مناسب است، اما در مورد هرچیز دیگر، بجز مقدار شاخصی که نمودار کرده اید، اطلاعاتی در اختیار نمی گذارد، ولی نمودار مستطیلی در مورد توزیع مقادیر در گروه ها مطالبی را بیان می کند. این نمودار به طور همزمان میانه، دامنه تغییرات بین چارکی و بیشترین و کمترین مقادیر یک گروه از نمونه ها را نشان می دهد [۴]. نمودار مستطیلی ابزاری مناسب برای شرح مکان و پراکندگی داده های مربوط به چند گروه مختلف است [۵]. (اما این نمودار برخلاف هیستوگرام، نمی تواند وجود چند گروه متفاوت در داده ها را نشان دهد).

در نمودار مستطیلی محور عمودی نمایش دهنده مقادیر متغیر و ویژگی اندازه گرفته شده است و محور افقی بیانگر گروه ها و فاکتورهای مورد نظر است. نحوه تشکیل این نمودار به شرح زیر است:

۱. ابتدا میانه^۲ و چارک های بالا^۳ و پایین^۴ محاسبه می شود (چارک پایین صدک ۲۵ام و چارک بالا صدک ۷۵ام است) [۵]. برای مثال داده های زیر را در نظر بگیرید که از کوچک به بزرگ مرتب شده اند:

۱۸، ۲۷، ۳۴، ۵۲، ۵۴، ۵۹، ۶۱، ۶۸، ۷۸، ۸۲، ۸۵، ۸۷، ۹۱، ۹۳، ۱۰۰

^۱box plot

^۲median

^۳upper quartile

^۴lower quartile

در این داده ها میانه ۶۸ است. سپس اعدادی که در سمت چپ میانه قرار دارند (یعنی ۶۱، ۵۹، ۵۴، ۵۲، ۳۴، ۲۷، ۱۸) در نظر گرفته شده و میانه این اعداد به عنوان چارک پایین معرفی می شود. در این مثال چارک پایین عدد ۵۲ می باشد. برای بدست آوردن چارک بالا باید میانه اعداد سمت راست میانه ی کل داده ها (یعنی ۱۰۰، ۹۳، ۹۱، ۸۷، ۸۵، ۸۲، ۷۸) استخراج شود. در اینجا چارک بالا ۸۷ است [۶].

۲. بر روی نمودار، مستطیلی کشیده می شود که بالا و پایین این مستطیل با چارک های بالا و پایین محدود می شود. یک خط نیز در محل میانه درون مستطیل رسم می شود. به این ترتیب $\frac{1}{4}$ ام از توزیع داده ها بین این خط و محدوده ی بالای مستطیل و $\frac{1}{4}$ ام از توزیع داده ها بین این خط و محدوده ی پایین مستطیل قرار می گیرد. این مستطیل نمایش دهنده محدوده ای است که ۵۰٪ میانی داده ها درون آن قرار می گیرد [۷]. کمیت دیگری که برای بیان این محدوده محاسبه می شود، محدوده میان چارکی یا IRQ است. این کمیت برابر با اختلاف بین چارک بالا و پایین می باشد. در مثال بالا IRQ برابر با $87 - 52 = 35$ می باشد [۶].

۳. سپس نقاط زیر محاسبه می شود:

$$L1 = 1.5 \times IRQ - \text{چارک پایین}$$

$$L2 = 3 \times IRQ - \text{چارک پایین}$$

$$U1 = 1.5 \times IRQ + \text{چارک بالا}$$

$$U2 = 3 \times IRQ + \text{چارک بالا}$$

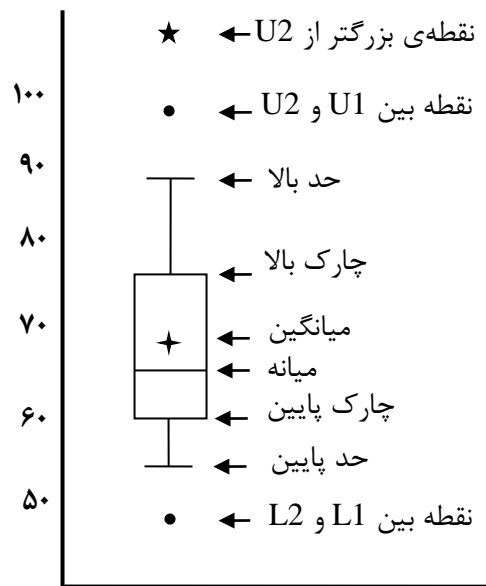
۴. در ادامه یک خط از چارک پایین تا کوچکترین نقطه که بزرگتر از $L1$ است رسم می شود. به همین صورت یک خط هم از چارک بالا تا بزرگترین نقطه که کوچکتر از $U1$ است، رسم می شود.

۵. نقاطی که بین $U1$ و $U2$ و یا $L1$ و $L2$ قرار دارند، با دایره های کوچک و نقاطی که بزرگتر از $U2$ و یا کوچکتر از $L2$ هستند، با دایره های بزرگتر یا علامت دیگر رسم می شود [۵].

مراحل ۳ تا ۵ برای رسم نمودار مستطیلی محصور شده است، برای رسم نمودار مستطیلی ساده بعد از مرحله ۳، یک خط از چارک بالا تا ماکسیمم نقاط و یک خط هم از چارک پایین تا مینیمم نقاط رسم می شود. در شکل ۳ نمونه ای از نمودار مستطیلی محصور شده آورده شده است.

^۱ interquartile range

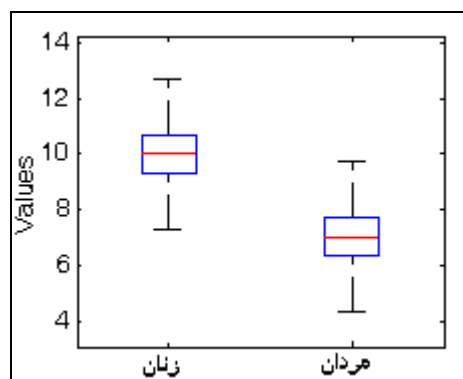
^۲Box plots with fences



شکل ۳: نمودار مستطیلی

به این ترتیب با رسم یک نمودار مستطیلی به ازای داده های اندازه گرفته شده از چند گروه مختلف می توان به صورت بصری تشخیص داد که آیا توزیع متغیر اندازه گرفته شده، در گروه های مختلف از نظر مکان و پراکندگی تفاوتی دارد یا خیر؟ همچنین می توان مشخص نمود که آیا توزیع یک متغیر، قبل و بعد از اعمال یک فاکتور تفاوت معنا داری دارد یا خیر؟ برای مثال به کمک این نمودار می توان به صورت دیداری تعیین نمود که آیا توزیع یک ویژگی قبل و بعد از القای هیپنوتیزم و یا در گروه های مختلف هیپنوتیزم پذیری تفاوت قابل توجهی دارد یا خیر؟ به طور کلی این نمودار می تواند پاسخگوی سوالات زیر باشد [۵]:

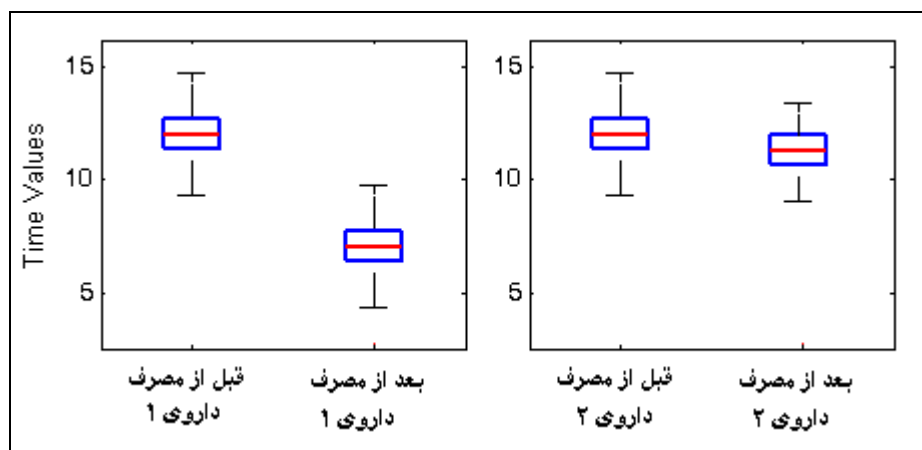
- آیا نحوه تغییر و پراکندگی نقاط مربوط به یک ویژگی (برای مثال میزان انرژی باند آلفا) اندازه گرفته شده ، در گروه های مختلف (برای مثال ۳ گروه هیپنوتیزم پذیری) تفاوت دارد؟ به عنوان نمونه در شکل زیر نمودار مستطیلی مربوط به زمان رسیدن به خط پایان در مسابقه دو در دو گروه زنان و مردان آورده شده است.



شکل ۴: نمودار مستطیلی حاصل از توزیع مدت زمان رسیدن به خط پایان در مسابقه دو در دو گروه مردان و زنان

با توجه به این شکل ملاحظه می شود که مدت زمان رسیدن به خط پایان در گروه مردان کمتر از زنان است.

- آیا اعمال یک فاکتور (برای مثال القای هیپنوتیزم) می تواند تغییر معناداری در یک ویژگی (برای مثال میزان انرژی باند آلفا) ایجاد کند؟ به عنوان نمونه در شکل بعدی تاثیر استفاده از دو داروی انرژی زای متفاوت، در مدت زمان رسیدن به خط پایان در مسابقه دو در یک مجموعه ۱۰۰۰۰ نفری از مردان آورده شده است.



شکل ۵: نمودار مستطیلی حاصل از توزیع مدت زمان رسیدن به خط پایان در مسابقه دو در یک مجموعه ۱۰۰۰۰ نفری از مردان، قبل و بعد از مصرف دو داروی متفاوت

با توجه به این شکل می توان گفت که مصرف داروی ۱ اثر قابل توجهی (کاهش) بر مدت زمان رسیدن به خط پایان داشته است. ولی داروی ۲ تاثیر قابل مشاهده ای در مدت زمان رسیدن به خط پایان نداشته است.

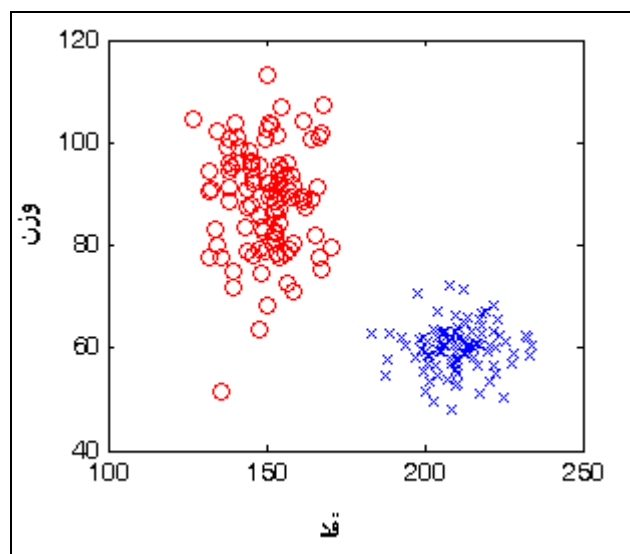
- آیا در دادگان، نقاط غیر معمولی^۱ وجود دارند که از محدوده داده ها فاصله داشته باشند؟

* در محیط مطلب برای رسم این نمودار از دستور **boxplot** استفاده می شود.

۳.۲. نمودار پراکندگی

زمانی که بخواهیم توزیع مقادیر مربوط به تعدادی نمونه را به ازای یک، دو یا سه ویژگی مختلف، مشاهده کنیم این نمودار می تواند مفید باشد. در این نمودار (برای مثال در مورد دو ویژگی) محور افقی مقادیر مربوط به یک ویژگی و محور عمودی مقادیر مربوط به ویژگی دیگر را نشان می دهد (در سه ویژگی، محور سوم مقادیر مربوط به ویژگی سوم را نشان می دهد). برای مثال در شکل زیر نمودار پراکندگی مربوط به وزن و قد (دو ویژگی متفاوت) دو گروه از افراد در دو نوع ورزش متفاوت را نشان می دهد.

^۱outliers



شکل ۶: نمودار پراکندگی مربوط به وزن و قد (دو ویژگی متفاوت) دو گروه از افراد در دو نوع ورزش متفاوت

با توجه به این نمودار می توان گفت که افراد در این دو نوع ورزش، از دو توزیع قد و وزن متفاوت برخوردار هستند. این نمودار زمانی که تعداد داده ها کم است نیز اطلاعات مفیدی را در اختیار می گذارد. به طور کلی این نمودار می تواند بیانگر موارد زیر باشد:

- نشان می دهد که آیا تجمع نمونه ها به سمت یک مقدار خاص است یا خیر؟
- وجود دسته ها و گروه های جداگانه در مقادیر نمونه ها را نشان می دهد. برای مثال زمانی که گروهی از نقاط در نمودار پراکندگی در یک سمت و گروهی دیگر در یک مکان دیگر تجمع می کنند، می تواند نشانگر وجود دو گروه متفاوت در داده ها است.
- آیا توزیع نقاط مربوط به یک یا چند ویژگی (حداکثر ۳ تا) ، در گروه های مختلف تفاوت دارد؟
- آیا اعمال یک فاکتور (برای مثال القای هیپنوتیزم) می تواند تغییر معناداری در یک ویژگی (برای مثال میزان انرژی باند آلفا) ایجاد کند؟
- آیا در دادگان، نقاط غیر معمولی وجود دارند که از محدوده داده ها فاصله داشته باشند.
- آیا یک ویژگی قدرت بیشتری نسبت به ویژگی های دیگر در جداسازی گروه ها دارد یا خیر؟ برای مثال در شکل ۸ مشاهده می شود که به ازای ویژگی قد دو گروه کاملاً از هم متمایز می شوند در صورتی که به ازای ویژگی وزن دو گروه کمی به هم همپوشانی دارند. به این ترتیب می توان نتیجه گرفت که ویژگی قد به تنهایی می تواند برای جداسازی دو گروه کافی باشد.

✱ در محیط مطلب برای رسم این نمودار از دستور *plot* و *plot3* یا *scatterplot* استفاده می شود.

۳. آزمون های معنی دار بودن

در بخش های قبلی با معرفی چند نمونه از نمودار های مختلف یک روش بصری برای تعیین و تشخیص معنادار بودن اختلافات ایجاد شده در گروه های مختلف توسط یک ویژگی و یا معنا دار بودن اختلاف موجود در یک ویژگی قبل و بعد از اعمال یک فاکتور، ارائه شد. اگر بخواهیم معنی دار بودن این اختلاف ها را نه به لحاظ بصری بلکه به صورت کمی بررسی نمائیم، از روشی به نام آزمون معنی دار بودن استفاده می شود.

انتخاب تصادفی نمونه ها از یک جامعه مفروض، نتایج حاصل از آزمون های معنی دار بودن را با ارزش می سازد [۸]. با توجه به توزیع داده ها، تعداد متغیرها و گروه ها، روشهای متعددی برای اجرای این آزمون ها ارائه شده است که در ادامه مختصرا به شرح تعدادی از آنها پرداخته خواهد شد.

۱.۳. فرض صفر^۱ و مخالف صفر^۲

قبل از به کار بردن هرگونه آزمون معنی دار بودن، باید فرضی مد نظر گرفته شود. اینگونه آزمونها معمولا فرض "هیچ اختلافی موجود نیست" را بیان می کنند که به نام فرض صفر نامیده شده و با H_0 نمایش داده می شود. برای مثال اگر بخواهیم بدانیم که آیا القای هیپنوتیزم در تغییر میزان باند آلفای امواج مغز تاثیر دارد یا خیر، فرض صفر را اینگونه تعیین می کنیم که "القای هیپنوتیزم هیچ اثری ندارد". مکمل فرض صفر را فرض مخالف صفر می نامند و آنرا با H_1 نشان می دهند. مشخص کردن فرض مخالف حائز اهمیت فراوان است. زیرا رد و یا قبول H_0 تنها زمانی مفهوم خواهد داشت که در مقابل یک فرض پایدار دیگری مورد آزمون قرار بگیرد [۸].

۲.۳. انواع خطاها در یک آزمون فرض

همانگونه که اشاره شد، نتیجه آزمون، در نهایت منجر به رد و یا قبول فرض H_0 می شود که در این تصمیم گیری خطایی نهفته است. در هر آزمون یکی از چهار تصمیم زیر ممکن است پیش بیاید:

(أ) H_0 رد می شود، موقعی که H_0 واقعا صحیح نباشد.

(ب) H_0 قبول می شود، موقعی که H_0 واقعا صحیح است.

(ج) H_0 رد می شود، موقعی که H_0 صحیح است.

(د) H_0 قبول می شود، موقعی که H_0 واقعا صحیح نباشد.

تصمیم های (أ و ب) تصمیم های درستی هستند، در حالی که تصمیم های (ج و د) غلط می باشد. خطای حاصل از تصمیم (ج) را خطای نوع اول و خطای حاصل از تصمیم (د) را خطای نوع دوم می نامند. اگر

$$\alpha = P[\text{خطای نوع اول}] = P[\text{رد } H_0 \text{ موقعی که صحیح است}]$$

$$\beta = P[\text{خطای نوع دوم}] = P[\text{قبول } H_0 \text{ موقعی که غلط است}]$$

باشد، α و β را اندازه خطای نوع اول و دوم می نامند. در یک آزمون خوب هم α و هم β باید مینیمم باشند. اما معمولا این امر میسر نیست. هر کوششی که منجر به مینیمم کردن α می شود، β را افزایش می دهد. در اغلب تصمیم گیری ها قبول یک فرض غلط، خیلی خطرناک تر از رد یک فرض صحیح است. یعنی خطای نوع دوم خیلی جدی تر از خطای نوع اول است. بنابراین معمولا α را ثابت گرفته و سعی در مینیمم کردن β دارند. مینیمم کردن β موجب ماکسیمم شدن $(1-\beta)$ می شود. $(1-\beta)$ را توان آزمون می نامند [۸].

^۱null hypothesis

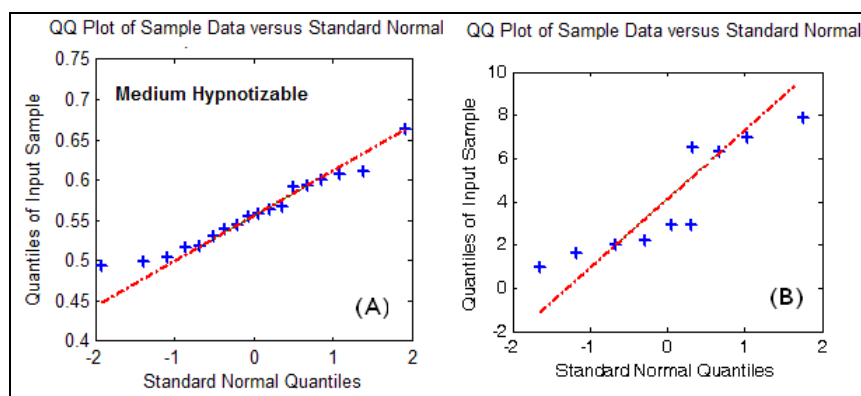
^۲alternative hypothesis

۳.۳. سطح معنی دار بودن

حداکثر اندازه α که حاضریم در مورد تصمیم گیری به آن اندازه خود را به خطر بیاندازیم، سطح معنی دار بودن نامیده می شود. اگر سطح معنی دار بودن، به اندازه α انتخاب شود، ایجاب می کند، در هر ۱۰۰ نمونه ای که انتخاب می شود، در $100 \times \alpha$ نمونه فرض صحیح رد شود. به عبارت دیگر $100(1-\alpha)$ درصد مطمئن هستیم که تصمیم ما در مورد رد فرض صفر صحیح است. موقعی که فرض صفر رد می شود، اطمینان معینی به تصمیم خود داریم که به سطح معنی دار بودن بستگی دارد. بنابراین در سطح معنی دار بودن " α "، درجه اطمینان برای تصمیم $(1-\alpha)$ خواهد بود که ضریب اطمینان نامیده می شود [۸].

۴.۳. ارزیابی فرض نرمال بودن

در بخشهای بعدی آزمون های مختلفی ارائه خواهد شد که همگی به فرض نرمال بودن توزیع متغیرها نیاز دارند. به همین دلیل قبل از به کارگیری این آزمون ها لازم است نرمال بودن متغیرها بررسی شود. از نمودار ویژه ای که نمودار Q-Q نامیده می شود، می توان برای ارزیابی فرض نرمال بودن استفاده کرد. برای هر مقدار داده ها، نمودار Q-Q مقدار مشاهده شده و مقدار منتظره (مربوط به حالتی است که داده های نمونه متعلق به توزیع نرمال است) را نشان می دهد. اگر داده ها متعلق به یک توزیع نرمال باشند، نقاط باید اطراف یک خط صاف جمع شوند. در شکل بعدی نمونه هایی از این نمودار برای دو توزیع نرمال و غیر نرمال آورده شده است.



شکل ۷: نمودار Q-Q به ازای دو توزیع (A) نرمال و (B) غیر نرمال

* در محیط مطلب برای رسم این نمودار از دستور *qqplot* استفاده می شود.

آزمون های آماری مرسوم نیز وجود دارد که به کمک آنها می توان این فرضیه صفر را آزمون نمود که آیا داده های مورد نظر، نمونه ای از جامعه نرمال می باشد. اگر سطح معنی داری مشاهده شده کوچک باشد، می توان به فرض نرمال بودن داده ها شک نمود [۴]. از جمله این آزمون ها می توان به آزمون های لیلی-فورس^۲، جاکر براؤ و کولموگروف-اسمیرنوف (KS)^۳ اشاره نمود. با اجرای این آزمون ها می توان اظهار داشت که با چه سطح معناداری می توان فرض صفر (نرمال بودن توزیع) را رد و یا قبول نمود.

^۱Quantile-quantile plot

Lilliefors

^۲Jarque-Bera

^۳Kolmogorov-Smirnov

* در محیط مطلب برای اجرای آزمون لیلی فورس از دستور *lillietest*، برای آزمون جاک برا از دستور *jbttest* و برای آزمون کولموگروف- اسمیرنوف از دستور *kstest2* استفاده می شود. آزمون KS نیاز به دانش قبلی نسبت به تابع چگالی احتمال داده ها دارد و اگر این تابع از روی داده ها تعیین شود، آزمون نمی تواند به خوبی عمل کند. آزمون جاک برا در مورد داده هایی با حجم کوچک نمی تواند به خوبی عمل کند ولی آزمون لیلی فورس برای داده هایی با حجم کوچک مناسب است [۹]. خروجی این آزمون ها در محیط مطلب صفر و یا یک است. صفر بودن خروجی نشان از نرمال بودن توزیع است. ولی خروجی یک نشان می دهد که فرض نرمال بودن توزیع باید رد شود. این آزمونها همچنین میزان سطح معناداری در رد یا قبول نرمال بودن توزیع را نیز ارائه می دهند. در نمونه های با حجم کوچک، ممکن است علی رغم نرمال نبودن فرض نرمال بودن را رد نمود. اگر فرض نرمال بودن مشکوک به نظر برسد و حجم نمونه کوچک باشد، می توان از تبدیل کردن مقادیر داده ها استفاده نمود تا توزیع بیشتر نرمال شود [۴].

۵.۳. تبدیلاتی برای نرمال کردن تقریبی

اگر نرمال بودن فرضی پا برجا نباشد، مرحله بعدی این است که از بررسی نرمال بودن صرف نظر کرده و با این فرض که داده ها نرمال هستند، عمل شود. اما این عمل در بسیاری از موارد باعث رسیدن به نتایج نادرست می شود. راه دوم این است که با در نظر گرفتن تبدیلات داده ها، داده های غیر نرمال "شبيه به نرمال" شوند. این تبدیلات در واقع چیزی جز بیان مجدد داده ها به واحد های مختلف نمی باشند. تبدیلات مناسب با بررسی های نظری یا خود داده ها یا هردو پیشنهاد می شوند. از طریق نظری، نشان داده شده است که داده های شمارشی را اغلب با گرفتن ریشه ی دوم آنها می توان بیشتر نرمال نمود البته توابع تبدیل متعددی برای این منظور ارائه شده است که برای دریافت جزئیات بیشتر آنها می توان به مرجع ۳ رجوع نمود. برای مطمئن شدن از این مسئله که آیا نرمال سازی رضایت بخش بوده است یا نه، می توان از نمودار Q-Q یا آزمون لیلی فورس که در بخش قبل معرفی شدند، استفاده نمود [۳]. اگر این عمل هم موفق نبود و داده ها از فرض نرمال بودن به شدت فاصله داشتند، می توان از آزمون هایی استفاده نمود که به فرض نرمال بودن توزیع داده ها نیازی ندارند [۴].

۶.۳. آنالیز واریانس یکطرفه^۱

هدف آنالیز واریانس یکطرفه یا ANOVA یکطرفه یافتن این حقیقت است که آیا داده های موجود در گروه های مختلف، که بر اساس یک ویژگی گروه بندی شده اند، دارای یک میانگین مشترک هستند یا خیر؟ به بیان دیگر این آنالیز تعیین می کند که آیا واقعا گروه ها، در مشخصه ی اندازه گرفته شده با هم تفاوت دارند یا خیر [۹]؟

فرضیه صفر در این آزمون این است که میانگین های گروه های مختلف یکی است، یعنی ویژگی اندازه گرفته شده تفاوت معناداری در گروه های مختلف ندارد. فرضیه مخالف صفر این است که تفاوت معناداری وجود دارد. فرضیه مخالف نمی گوید که کدام گروه با بقیه فرق دارد، بلکه صرفا می گوید میانگین متغیر مورد نظر در گروه ها یکسان نمی باشد و حداقل یکی از گروه ها با بقیه فرق می کند.

^۱One-Way ANalysis Of VAriance

این آزمون پراکندگی مقادیر نمونه را مورد بررسی قرار می دهد و بررسی می کند که چقدر مشاهدات درون هر گروه و همچنین میانگین های گروهی دارای پراکندگی هستند. براساس برآورد این دو پراکندگی می توان در مورد میانگین های گروه های مختلف نتیجه گیری کرد. اگر پراکندگی میانگین های گروهی بیشتر از پراکندگی موجود در مشاهدات هر گروه تغییر کند، می توان نتیجه گرفت که میانگین های گروه های مختلف نسبت به هم تفاوت معناداری دارند. این آنالیز به فرضهای زیر نیاز دارد:

- از هر گروه نمونه های تصادفی مستقل گرفته شده باشد. به این معنی که رابطه ای بین مشاهده های گروه های مختلف یا بین مشاهده های یک گروه وجود نداشته باشد. برای مثال اگر چهار درمان مختلف به هریک از نمونه ها تجویز شده باشد، نمی توان برای آنالیز داده ها (بررسی اثر چهار درمان) از آنالیز واریانس یکطرفه استفاده نمود. نمونه ها در این مثال مستقل نیستند زیرا از هریک از افراد مشاهداتی در هر گروه وجود دارد (یک نفر در بیشتر از یک گروه شرکت داشته است).
- توزیع گروه ها نرمال باشد. در عمل آنالیز واریانس خیلی به فرض نرمال بودن وابسته نمی باشد مگر اینکه داده ها خیلی از نرمال بودن فاصله گرفته باشند. اگر حجم نمونه در گروه ها کوچک باشد، بایستی در مورد اثر مشاهدات غیر معمول که می توانند اثرات مهمی روی انحراف معیار و میانگین بگذارند، دقت داشت. برای اطمینان می توان بدون نقطه غیرمعمول موجود در داده ها آنالیز را مجدداً اجرا کرد تا مطمئن شد که همان نتایج قبلی بدست می آید یا خیر؟
- واریانسهای گروه ها برابر باشد. فرض برابری واریانس را از طریق بررسی پراکندگی مشاهدات در نمودار مستطیلی می توان بررسی نمود. علاوه بر نمودار مستطیلی، آزمون دیگری نیز برای بررسی برابری واریانسها، به نام *levene* وجود دارد. در عمل اگر تعداد نمونه ها در هر یک از گروه ها شبیه به هم باشد فرض برابری واریانسها چندان مهم نمی باشد [۴].

علاوه بر آنالیز واریانس یکطرفه، آنالیزهای واریانس دو و یا N طرفه نیز وجود دارد. زمانی که داده ها بر یک اساس به چند گروه طبقه بندی شده باشند و بخواهیم بدانیم که آیا یک ویژگی توانایی ایجاد تمایز معنادار بین این گروه ها را دارد یا خیر، از آنالیز واریانس یکطرفه استفاده می شود. ولی زمانی که داده ها بر دو یا چند اساس مختلف به چند گروه طبقه بندی شوند و بخواهیم بدانیم که آیا یک ویژگی توانایی ایجاد تمایز معنادار بین این گروه ها (بر هر دو یا چند اساس) را دارد یا خیر، از آنالیز واریانس دو یا چند متغیره استفاده می شود. برای مثال فرض کنید که داده ها بر اساس جنسیت، به دو گروه زن و مرد تقسیم شده باشند و بخواهیم بدانیم که آیا، ویژگی مدت زمان استفاده از اینترنت می تواند بین این دو گروه تمایز معنادار ایجاد کند یا خیر، در این صورت می توان از آنالیز واریانس یکطرفه استفاده نمود. ولی اگر داده ها بر اساس جنسیت، به دو گروه زن و مرد و بر اساس سن، به دو گروه پیر و جوان تقسیم شده باشند و بخواهیم بدانیم که آیا، ویژگی مدت زمان استفاده از اینترنت می تواند بین گروه های زن و مرد و پیر و جوان تمایز معنادار ایجاد کند یا خیر، از آنالیز واریانس دو طرفه استفاده می شود.

در آنالیز واریانس، پراکندگی مشاهده شده در نمونه ها به دو بخش تقسیم می شود: پراکندگی مشاهدات درون یک گروه حول میانگین گروه و پراکندگی بین میانگین گروه ها. این برآوردهای پراکندگی در جدولی به

نام جدول آنالیز واریانس (جدول ۱) نشان داده می شود که به کمک این جدول می توان گفت که آیا شواهد کافی مبتنی بر رد کردن فرضیه صفر وجود دارد یاخیر؟

جدول ۱: جدول آنالیز واریانس یکطرفه (ردیف Columns: مقادیر بین گروهی، ردیف Error: مقادیر درون گروهی، ستون SS: مجموع مربعات پراکندگی ها، df: درجه آزادی، MS: میانگین مربعات پراکندگی ها، F: نسبت MS ها، Prob>F: سطح معنی داری (p-value)، افزایش F باعث کاهش سطح معنی داری می شود) [۹]

ANOVA Table					
Source	SS	df	MS	F	Prob>F
Groups	174.438	2	87.219	87.22	5.40124e-013
Error	29	29	1		
Total	203.438	31			

مقدار دو پراکندگی در ستونی به نام میانگین مربعات (MS) نشان داده می شود. نسبت مربع میانگین بین گروهی به مربع میانگین درون گروهی در ستونی به نام F ارائه می شود. اگر فرضیه صفر صحیح باشد، انتظار می رود که این نسبت نزدیک به یک باشد. با نگاه کردن به سطح معنی داری مشاهده شده در ستون Prob>F می توان گفت که آیا نسبت F مشاهده شده به اندازه کافی بزرگ می باشد که بتوان فرضیه صفر را رد نمود. با توجه به جدول ۱ ملاحظه می شود در صورتی که فرضیه صفر صحیح باشد، احتمال بدست آوردن نسبت F برابر یا بزرگتر از ۱۹,۱۸ حدود 1.28×10^{-6} می باشد. به این ترتیب هنگامی که فرضیه صفر صحیح است، تنها ۱,۲۶ در ۱۰۰۰۰۰۰ انتظار می رود که نسبتی برابر با مقدار مشاهده شده یا بزرگتر را بدست آورد، بنابراین می توان فرضیه صفر را رد نمود [۴].

برای محاسبه مقادیر ذکر شده در جدول به صورت زیر عمل می شود:

محاسبه پراکندگی درون گروهی: محاسبه این پراکندگی دارای سه مرحله است: ۱- ابتدا باید مقداری را محاسبه نمود که مجموع مربعات درون گروهی^۱ نامیده می شود. ابتدا واریانسهای مربوط به مشاهدات موجود در هر گروه محاسبه شده و هر کدام از واریانسها در تعداد نمونه های گروه مربوطه منهای یک ضرب می شود. در انتها مقادیر بدست آمده در تمامی گروه ها با هم جمع می شود. این عدد در ستون SS در ردیف دوم قابل مشاهده است. ۲- در مرحله بعدی باید درجه آزادی را محاسبه نمود. برای محاسبه درجه آزادی درون گروهی، در هر گروه تعداد نمونه ها منهای یک شده و با هم جمع می شوند. این عدد در ستونی به نام df در ردیف دوم مشاهده می شود. ۳- در آخر مجموع مربعات بدست آمده بر درجه آزادی تقسیم می شود تا مربع میانگین (MS) بدست آید. این عدد برآوردی از متوسط پراکندگی ها در گروه ها است. این عدد در حقیقت چیزی نیست جز متوسط واریانس تمامی گروه ها با توجه به این واقعیت که تعداد مشاهدات در گروه ها با هم فرق می کند.

محاسبه پراکندگی بین گروهی: برای محاسبه این پراکندگی نیز سه مرحله وجود دارد. ۱- ابتدا مجموع مربعات بین گروهی^۲ محاسبه می شود. میانگین کل (میانگین تمامی مشاهدات)، از میانگین هر گروه کم می شود. سپس هر کدام از این تفاوت ها به توان دو رسیده و در تعداد مشاهدات مربوط به هر گروه ضرب می

^۱Mean Square

^۲Within-groups sum of square

^۳between-groups sum of square

شود و در نهایت تمامی مقادیر بدست آمده با هم جمع می شود. این عدد در ستوت SS در ردیف اول مشاهده می شود. ۲- در مرحله بعد باید درجه آزادی مجموع مربعات بین گروهی را محاسبه نمود که عبارت است از تعداد گروه ها منهای یک. این عدد در ستون df، ردیف اول نشان داده می شود. ۳- در مرحله آخر با تقسیم مجموع مربعات بدست آمده بر درجه آزادی مقدار مربع میانگین بین گروهی محاسبه می شود که در ستون MS ردیف اول قابل مشاهده است.

(باید به خاطر داشت که مربع میانگین درون گروهی براساس پراکندگی مشاهدات هر گروه محاسبه می شود و مربع میانگین بین گروهی بر این اساس است که میانگین گروه ها چقدر با هم تفاوت می کند)

محاسبه نسبت F و سطح معنی داری: حال که دو پراکندگی مورد نظر محاسبه شده است، نسبت F را می توان به صورت زیر بدست آورد:

$$F = \frac{\text{پراکندگی بین گروهی (MS)}}{\text{پراکندگی درون گروهی}}$$

از طریق مقایسه نسبت F محاسبه شده با مقدار توزیع F سطح معنی داری مشاهده شده محاسبه می شود. سطح معنی داری مشاهده شده هم به نسبت F و هم به درجه آزادی دو مربع میانگین وابسته است [۴]. آنالیز واریانس یکطرفه یک مثال ساده از یک مدل خطی است که مدل آن به صورت زیر قابل بیان است:

$$y_{ij} = \alpha_{.j} + \varepsilon_{ij}$$

که در این رابطه y_{ij} ماتریس مشاهدات است که هر ستون آن بیانگر یک گروه مختلف است؛ $\alpha_{.j}$ ماتریسی است که ستونهای آن میانگین داده های مربوط به هر گروه است (اندیس "j" به این معناست که α به تمامی ردیف های ستون j ام اعمال می شود)؛ ε_{ij} ماتریسی از اختلالات تصادفی است.

به بیان دیگر این مدل فرض می کند که ستون های y (داده های هر گروه)، از یک عدد ثابت (میانگین) به علاوه یک اختلال تصادفی (پراکندگی های موجود در داده ها حول میانگین) تشکیل شده اند. هدف از آنالیز این است که آیا این اعداد ثابت واقعا با هم یکی هستند یا خیر [۹]؟

* در محیط مطلب برای اجرای این آزمون از دستور **anova1** استفاده می شود.

آزمون دیگری به نام آزمون t با نمونه های مستقل نیز وجود دارد که می تواند برای بررسی توانایی یک ویژگی در ایجاد تفاوت معنادار بین دو گروه با نمونه های مستقل به کار رود. تفاوت این آزمون با آزمون ANOVA در این است که آزمون t فقط تفاوت های بین دو گروه را بررسی می کند. اگر دو گروه مستقل وجود داشته باشد، نتایجی که از آزمون ANOVA بدست می آید، با نتایج آزمون t با واریانس های برابر یکسان است [۴].

* در محیط مطلب برای اجرای آزمون t با نمونه های مستقل از دستور **ttest2** استفاده می شود.

¹Independent sample t test

۷.۳. روشهای مقایسه چندگانه^۱

آزمونهای آنالیز واریانس تنها می توانند بگویند که از نظر آماری ظاهراً غیر محتمل است که میانگین های تمامی گروه ها با هم برابر باشد. ولی این آزمونها نمی توانند نشان بدهند که کدام گروه ها با بقیه متفاوت هستند. معمولاً هنگامی که یک فرضیه صفر رد می شود، انتظار می رود که محل اختلاف ها نیز مشخص شود. برای این کار از روشهای مقایسه چند گانه استفاده می شود.

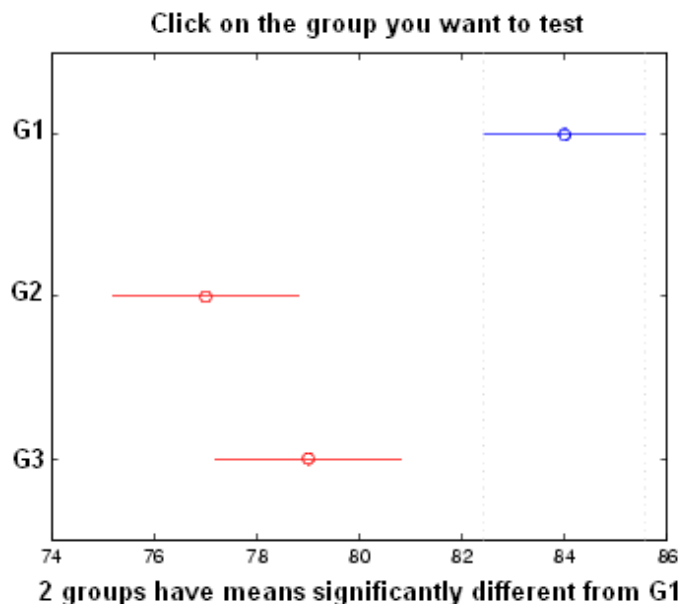
* در محیط مطلب برای اجرای این آزمون از دستور *multcompare* استفاده می شود.

ورودی این آزمون شامل اطلاعاتی است که در جدول آنالیز واریانس یکطرفه وجود دارد و خروجی آن ماتریسی است که هر ردیف آن شامل اعدادی مانند اعداد زیر است.

۲,۰۰۰۰ ۵,۰۰۰۰ ۱,۹۴۴۲ ۸,۲۲۰۶ ۱۴,۴۹۷۱

این اعداد نشان می دهد که میانگین توزیع در گروه ۲ منهای گروه ۵ حدود ۸,۲۲۰۶ است و حداقل اختلاف بین مقادیر این دو گروه برابر با ۱,۹۴۴۲ و حداکثر ۱۴,۴۹۷۱ است. در این مثال این فاصله عدد صفر را شامل نمی شود. در نتیجه با اطمینان ۹۵٪ می توان گفت که تفاوت بین این دو گروه معنادار است. اگر این فاصله صفر را شامل شود، می توان گفت که در بین این دو گروه تفاوت معناداری وجود ندارد.

با اجرای این آزمون در محیط مطلب علاوه بر اطلاعات بالا نموداری به صورت شکل زیر نیز نمایش داده می شود، که در واقع نموداری از توزیع مقادیر هر گروه است. با انتخاب هر گروه می توان مشاهده نمود که آن گروه با کدامیک از گروه ها تفاوت معنادار دارد.



شکل ۸: نمودار حاصل از اجرای آزمون مقایسه چندگانه در محیط مطلب

با توجه به شکل بالا مشاهده می شود که تفاوت معنادار موجود بین گروه G1 با دو گروه دیگر است. البته به غیر از این روش می توان از نمودار مستطیلی و یا نمودار پراکندگی نیز استفاده نمود.

^۱Multiple comparison

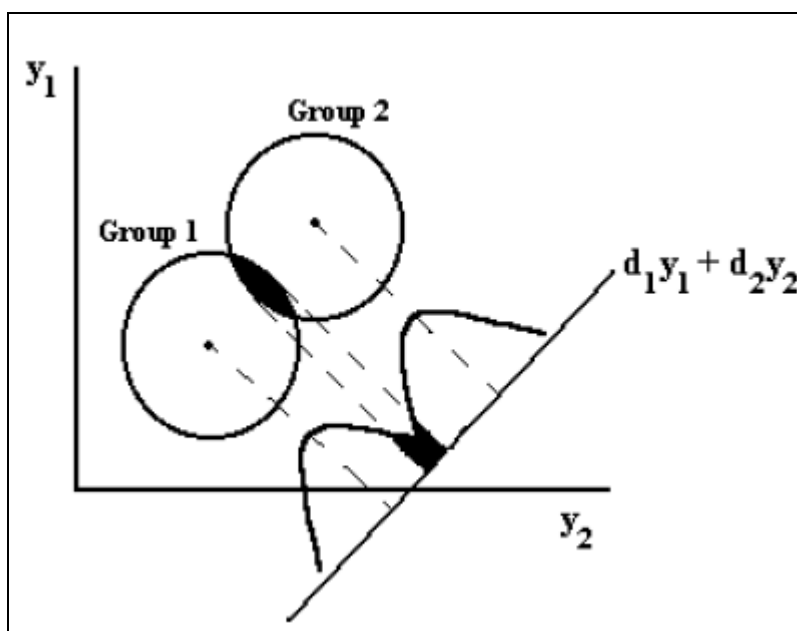
۴. آنالیز واریانس چند متغیره یکطرفه^۱

در بخش ۶،۳ آنالیز واریانس یکطرفه یا ANOVA، مورد بررسی قرار گرفته شد. این آنالیز مشخص می نمود که آیا میانگین یک ویژگی در بین گروه های مختلف تفاوت معناداری دارد یا خیر؟ ممکن است بخواهیم اثر چند ویژگی را در ایجاد تفاوت بین گروه های مختلف مورد بررسی قرار دهیم، برای انجام این کار می توانیم از آنالیز واریانس چند متغیره ی یکطرفه استفاده کنیم.

گاهی ممکن است یک ویژگی به تنهایی نتواند تمایز معناداری را بین گروه های مورد نظر ایجاد کند. ولی ممکن است که استفاده همزمان از دو و یا چند ویژگی بتواند باعث ایجاد تمایز معنادار بین گروه های هیپنوتیزم پذیری بشود. بررسی این مسئله به کمک آنالیز واریانس چند متغیره یا MANOVA، قابل انجام است.

البته باید توجه داشت که آزمون MANOVA زمانی که تعداد ویژگی ها خیلی بیشتر از تعداد نمونه های درون گروه هاست، نمی تواند جواب بدهد. اساس کار این آزمون نیز مانند آزمون ANOVA، بر اساس مقایسه پراکندگی بین گروهی به پراکندگی درون گروهی است (محاسبه نسبت F برای چند متغیر تعمیم داده میشود).

به کمک آزمون MANOVA همچنین می توان یک ترکیب خطی از چند ویژگی را ایجاد نمود که بیشترین جداسازی را در گروه های مختلف ایجاد کند. نتیجه ترکیب بدست آمده می تواند به عنوان یک ویژگی جدید در نظر گرفته شود که بیشترین تفاوت معنادار را آنالیز واریانس یکطرفه (تک متغیره) یا ANOVA ایجاد می کند [۱۰ و ۹]. برای مثال در شکل زیر یک ترکیب خطی از دو ویژگی y_1 و y_2 ، برای جدا سازی دو گروه ایجاد شده است.



شکل ۹: ایجاد یک ترکیب خطی از دو ویژگی جهت جدا سازی دو گروه مختلف

^۱One-Way Multivariate ANalysis Of VAriance

محاسبه ضرایب این ترکیب خطی جدید به نحوی انجام می شود که به ازای ترکیب خطی جدید بدست آمده، پراکندگی درون گروهی مینیمم و پراکندگی بین گروهی ماکسیمم شود، به عبارت دیگر نسبت F ماکسیمم شود. این روش مشابه با " آنالیز تابع متمایز کننده" یا^۲ LDA چند کلاسه عمل می کند. با در نظر گرفتن ترکیب خطی مربوط به ویژگی های مورد نظر نسبت F رابطه ای به صورت رابطه زیر تشکیل می دهد.

$$F = \frac{\vec{W}^T \Sigma_b \vec{W}}{\vec{W}^T \Sigma \vec{W}}$$

که در این رابطه، Σ_b پراکندگی بین گروهی، Σ پراکندگی درون گروهی و بردار W ، دربردارنده ضرایب ترکیب خطی مورد نظر است. هدف پیدا نمودن بردار ضرایب است به نحوی که نسبت F ماکسیمم شود. به این ترتیب اگر بردار W یکی از بردارهای ویژه^۳ $\Sigma^{-1} \Sigma_b$ باشد، تمایز بین گروه ها برابر با مقدار ویژه^۴ متناظر با آن خواهد بود.

* در محیط مطلب برای اجرای این آزمون از دستور *manova1* استفاده می شود. اجرای این آزمون در این محیط چند خروجی مختلف تولید میکند که از بین آنها اولین پارامتر، d ، تخمینی از بعد میانگین ویژگی جدید بدست آمده در گروه های مختلف است. اگر میانگین ها مشابه هم باشند، یعنی گروه ها همدیگر را بپوشانند، آنگاه بعد بدست آمده صفر خواهد بود به بیانی، میانگین تمام گروه ها بر روی یک نقطه است. اگر میانگین ها متفاوت باشد ولی در امتداد یک خط قرار بگیرند، بعد یک خواهد بود. اگر بعد دو باشد یعنی میانگین ها نه در امتداد یک خط، بلکه بر روی یک صفحه قرار می گیرند. بعد دو بزرگترین بعد ممکن برای مسئله ای با سه گروه است. دومین پارامتر خروجی، p ، برداری از مقادیر p -value یا سطوح معناداری است. اولین مقدار از بردار p ، مقدار سطح معناداری در صفر بودن بعد، دومین مقدار، سطح معناداری در یک بودن بعد را نشان می دهد و به همین ترتیب برای مقادیر بعدی. سومین خروجی، $stats$ ، خود متشکل از چند فیلد است: فیلدهای W ، B و T ماتریسهایی هستند که مشابه با آنالیز واریانس یک طرفه ANOVA به ترتیب شامل، مجموع مربعات درون گروهی، بین گروهی و مجموع مربعات کل می باشند. سه پارامتر بعدی dfW ، dfT ، dfB درجات آزادی مربوط به این سه ماتریس هستند. فیلد $eigenvec$ ماتریسی است که ضرایب ترکیبات خطی ویژگی های اصلی را در بر دارد و اولین ستون آن مربوط به ضرایبی است که ترکیب خطی متناظر با آن بیشترین تفاوت معنادار را بین گروه ها ایجاد می کند. $eigenval$ برداری است که نسبت واریانس بین گروهی به درون گروهی مربوط به ترکیب خطی متناظر با آن را در بر دارد.

۱.۴. آزمون t با نمونه های جفت^۵

از این آزمون جهت آنالیز نتایج مطالعاتی استفاده می شود که هر فرد دوبار در دو وضعیت متفاوت مورد مشاهده قرار می گیرد یا در مطالعاتی که در آن جفت هایی از افراد (یا اندازه گیری های جفت) وجود دارد که به طریقی به هم وابسته هستند. یکی از این نوع مطالعات طرح "قبل و بعد" است. برای مثال اگر بخواهیم

^۱Discriminant Function Analysis

^۲Linear Discriminant Analysis

^۳eigenvector

^۴eigenvalue

^۵Paired sample t test

بدانیم که آیا دویدن باعث افزایش سطح بتآندروفین در خون می شود یا خیر، مجبور هستیم سطح این متغیر را در هر سوژه دوبار اندازه گیری کنیم، یک بار قبل از دویدن و یک بار بعد از دویدن، این مطالعه مثال خوبی از یک طرح جفت است (به عنوان مثالی دیگر می توان به مطالعه اثر القای هیپنوتیزم در سطح انرژی امواج دلتا در سیگنال مغزی، اشاره نمود).

مزیت طرح جفت آن است که سبب می شود که اگر اختلافی وجود داشته باشد، راحت تر کشف شود. مانند آزمونهای قبلی در آزمون t با نمونه های جفت نیز لازم است که داده های اندازه گیری شده (قبل از القای هیپنوتیزم و بعد از القای هیپنوتیزم) از یک جامعه نرمال فاصله زیادی نداشته باشند. همچنین باید اندازه گیری متغیر مورد نظر از یک گروه از افراد باشد، اگر مقادیر بدست آمده از دو گروه متفاوت از افراد باشد (یک گروه قبل از القای هیپنوتیزم و یک گروه بعد از القای هیپنوتیزم)، مقداری از اختلاف مشاهده شده بین دو میانگین می تواند ناشی از اختلاف ذاتی بین افراد موجود در دو گروه باشد. برای مثال افرادی که قبل از القای هیپنوتیزم اندازه گیری شده اند، ممکن است به طور طبیعی سطح انرژی امواج دلتا در آنها پایین تر و یا بالاتر باشد.

فرضیه صفر در یک طرح جفت این است که اختلافی بین مقادیر متوسط دو عضو یک جفت وجود ندارد (به بیان دیگر فاکتور مورد نظر اثر معناداری بر ویژگی در نظر گرفته شده ندارد). فرض مخالف این است که در مقادیر متوسط اختلافی وجود دارد [۴].

برای اجرای این آزمون با در نظر گرفتن وضعیت زیر در مورد داده ها:

$$1. \text{ نمونه ها دارای حجم مساوی باشند یعنی } n_1 = n_2 = n$$

۲. مشاهدات نمونه ای $(X_1, X_2, \dots, X_n), (Y_1, Y_2, \dots, Y_n)$ کاملاً مستقل بوده ولی به صورت زوج وابسته باشند، یعنی زوج مشاهدات $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ به ترتیب متناظر با اولین، دومین و... آخرین سوژه باشند.

تحت فرض $H_0: \mu_x = \mu_y$ ، آماره آزمون با استفاده از رابطه زیر محاسبه می شود:

$$t = \frac{\bar{d}}{S / \sqrt{n}} = \frac{\bar{d}}{\sqrt{S^2 / n}} \approx t_{n-1}$$

$$\bar{d} = \frac{\sum d}{n}, \quad d = x - y$$

$$S^2 = \frac{1}{n-1} \sum (d - \bar{d})^2 = \frac{1}{n-1} \left[\sum d^2 - \frac{(\sum d)^2}{n} \right]$$

سپس مقدار t بدست آمده با مقدار t جدول t استیودنت، به ازای درجه آزادی $n-1$ و سطح معنی داری 0.01 و یا 0.05 مقایسه می شود، اگر t محاسبه شده کمتر باشد، نتیجه می شود که در سطح 1% یا 5% درصد، اختلاف معنی دار نیست و می توان فرض صفر را قبول کرد [۸].

* در محیط مطلب برای اجرای این آزمون از دستور **ttest** استفاده می شود.

همانطوری که اشاره شد، آزمون t با نمونه های جفت، زمانی به کار می رود که هدف یافتن تفاوت معنادار بین ویژگی اندازه گرفته شده از یک گروه از افراد در قبل و بعد از اعمال یک فاکتور باشد. در این موارد ممکن است این اشتباه به وجود آید که، داده های مربوط به قبل از اعمال فاکتور به عنوان یک گروه (کلاس) و داده های مربوط به بعد از اعمال فاکتور به عنوان گروه (کلاس) دوم در نظر گرفته شود. سپس برای تشخیص

وجود تمایز معنادار بین این دو گروه از آزمون‌های آنالیز واریانس (ANOVA) و یا آزمون t با نمونه‌های مستقل (ttest2) استفاده شود. اگر با چنین داده‌هایی از این دو آزمون استفاده شود، ممکن است برحسب اتفاق جواب صحیح هم بدست آید، ولی به طور کلی نباید به نتایج اطمینان داشت. زیرا یک شرط لازم در این دو آزمون (ANOVA) (MANOVA) و (ttest2) مستقل بودن داده‌های مربوط به گروه‌های مختلف است. در صورتی که در این موارد، داده‌های مربوط دو کلاس فرض شده از یک گروه از افراد ثبت شده است و در نتیجه داده‌های دو کلاس نسبت به هم مستقل نیست.

۵. منحنی ROC^۱ و پارامتر AUC^۲

منحنی ROC تکنیکی جهت بررسی و مشاهده نتایج طبقه‌بندی گروه‌ها با ویژگی‌ها و طبقه‌بندی‌کننده‌های مختلف است. بعضی از طبقه‌بندی‌کننده‌ها دارای خروجی پیوسته هستند (تخمینی از احتمال عضویت هر نمونه به کلاسها) که با به کارگیری یک حد آستانه عضویت به یک کلاس را پیشبینی می‌کنند. بعضی از مدلها در خروجی یک عدد گسسته (برچسب کلاس) تولید می‌کنند، که فقط پیشبینی می‌کنند نمونه به کدام کلاس متعلق است. معمولاً منحنی ROC در مورد طبقه‌بندی و جدا سازی دو گروه مطرح می‌شود، بنابراین در اینجا، ابتدا بحث‌های مربوطه براساس مسئله طبقه‌بندی و جداسازی دو گروه مطرح می‌شود و در ادامه نتایج بدست آمده به طبقه‌بندی چند کلاسه بسط داده می‌شود. با وجود دادن یک مشاهده به یک طبقه‌بندی‌کننده، چهار حالت ممکن است اتفاق بیافتد (در مسائلی با دو گروه). این چهار حالت در شکل زیر خلاصه شده‌اند.

		True class	
		p	n
Hypothesized class	Y	True Positives	False Positives
	N	False Negatives	True Negatives
Column totals:		P	N

$$fp\ rate = \frac{FP}{N} \quad tp\ rate = \frac{TP}{P}$$

$$sensitivity = tp\ rate$$

$$specificity = 1 - fp\ rate$$

$$precision = \frac{TP}{TP + FP}$$

$$accuracy = \frac{TP + TN}{P + N}$$

$$Fmeasure = \frac{2}{1/precision + 1/sensitivity}$$

شکل ۱۰: جدول احتمال به همراه معیارهای ارزیابی مختلف که به کمک این جدول محاسبه شده‌اند [۱۰].

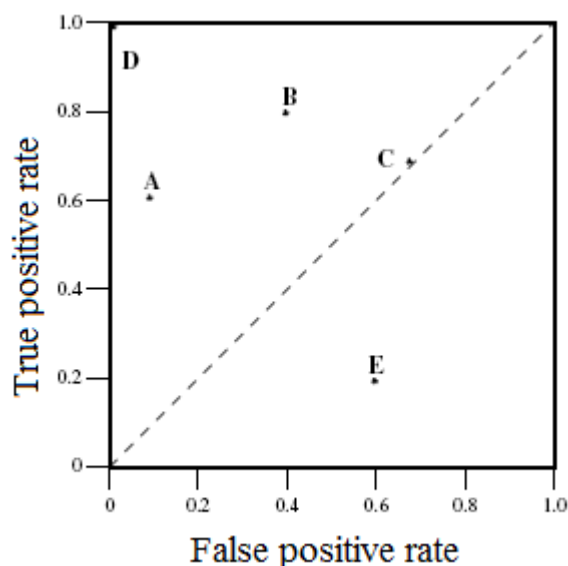
منحنی ROC در مورد مسائلی با دو گروه، یک منحنی دو بعدی است که در آن نرخ tp بر روی محور Y و نرخ fp بر محور X رسم می‌شود. نمودار ROC مصالحه‌ی نسبی بین سود (tp) و هزینه (fp) را شرح می‌دهد. شکل ۱۱ نمودار ROC را برای ۵ طبقه‌بندی‌کننده مختلف که با حروف A تا E بر چسب خورده‌اند نشان می‌دهد.

^۱Receiver Operating Characteristic

^۲Area Under Curve

یک طبقه بندی کننده ی گسسته، طبقه بندی کننده ای است که خروجی آن فقط برچسب کلاس را می دهد. هر طبقه بندی کننده ی گسسته ای یک جفت (fp, tp) تولید می کند که متناظر با یک نقطه بر روی منحنی ROC است. طبقه بندی کننده های شکل ۱۱ همگی گسسته هستند.

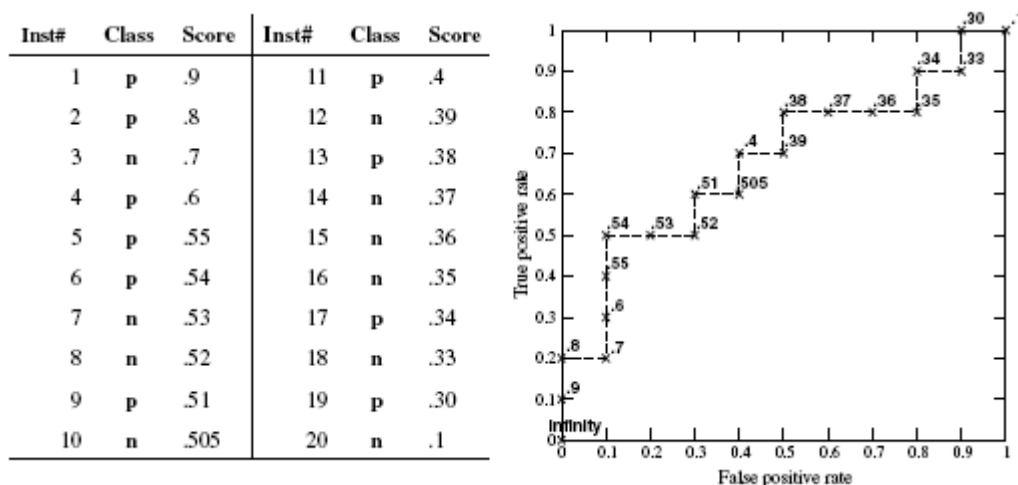
در شکل ۱۱ پایین ترین نقطه ی سمت چپ $(0,0)$ نشان دهنده ی طبقه بندی کننده ای است که هیچگاه نتیجه ی مثبت بودن نمونه را نمی دهد. به این ترتیب این چنین طبقه بندی کننده ای مرتکب خطای به اشتباه مثبت (fp) نمی شود. ولی در عوض اگر نمونه واقعا مثبت باشد (tp) بازهم نتیجه ای نمی دهد. حالت عکس در مورد طبقه بندی کننده ای اتفاق می افتد که در نمودار ROC بر روی نقطه ی $(1,1)$ نمایش داده می شود. نقطه ی موجود در $(1,0)$ یک طبقه بندی بی عیب و کامل را نشان می دهد. بنابراین طبقه بندی کننده ی D یک طبقه بندی کننده ی کامل است. به این ترتیب می توان گفت یک نقطه بر روی منحنی ROC بهتر از دیگری است اگر در سمت بالا و یا چپ (tp) بیشتر یا (fp) کمتر یا هر دو نقطه ی اول قرار داشته باشد.



شکل ۱۱: یک نمونه ی ابتدایی از نمودار ROC که نتایج حاصل از ۵ طبقه بندی کننده ی گسسته بر روی آن نمایش داده شده است [۱۰].

طبقه بندی کننده هایی که در سمت چپ نمودار ROC نزدیک محور X ظاهر می شوند، به عنوان طبقه بندی کننده های "محافظه کار" شناخته می شوند. این نوع طبقه بندی کننده ها تنها زمانی که اطمینان کافی وجود داشته باشد، نتیجه ی مثبت بودن نمونه را می دهند و به این ترتیب خطای به اشتباه مثبت (fp) خیلی کمی دارند، ولی متناظرا نرخ tp آنها نیز کم می باشد. طبقه بندی کننده هایی که در سمت بالا و راست منحنی ROC قرار می گیرند، به عنوان طبقه بندی کننده های آزاد یا "لیبرال" شناخته می شوند. این نوع طبقه بندی کننده حتی زمانی که اطمینان کمی نسبت به مثبت بودن نمونه داشته باشد، نتیجه ی مثبت می دهد و به این ترتیب ممکن است خطای fp بالایی داشته باشند. در شکل A محافظه کارتر از B است. از آنجایی که در دنیای واقعی تعداد نمونه هایی که واقعا منفی هستند، بیشتر است بنابراین نقاطی که سمت چپ نمودار ROC قرار دارند، بیشتر مورد توجه قرار می گیرند.

طبقه بندی کننده هایی که بر روی خط قطری $y=x$ قرار می گیرند، طبقه بندی کننده ای هستند که به طور تصادفی مثبت بودن نمونه را در ۵۰ درصد زمان، نتیجه می دهد. به این ترتیب انتظار داریم نتیجه ی بدست آمده از منفی بودن و یا مثبت بودن نمونه، ۵۰٪ صحیح باشد که منجر به نقطه ی (۰,۵ و ۰,۵) بر روی نمودار ROC می شود. اگر یک طبقه بندی کننده ۹۰٪ زمانها نمونه را مثبت تشخیص بدهد، انتظار داریم که نتیجه ی بدست آمده ۹۰٪ به درستی مثبت (tp) و همچنین ۹۰٪ به اشتباه مثبت (fp) باشد و نقطه ی (۰,۹ و ۰,۹) را روی نمودار ROC نمایش می دهد. به این ترتیب نقاط حاصل از طبقه بندی کننده های تصادفی در اطراف خط قطری قرار می گیرند و مکان آنها بر روی این خط با توجه به تناوب و نرخ تشخیص مثبت بودن نمونه تعیین می شود. اگر یک طبقه بندی کننده بخواهد از سمت این خط قطری به سمت مثلث بالایی نمودار حرکت کند، باید به جای حدس تصادفی از اطلاعات موجود در داده ها استفاده نماید. طبقه بندی کننده هایی که در مثلث پایینی قرار می گیرند، بدتر از یک طبقه بندی کننده تصادفی عمل می کنند. اما بعضی از طبقه بندی کننده ها به جای تولید یک خروجی گسسته (برچسب کلاس) یک درجه به هر ورودی نسبت می دهد، سپس یک سطح آستانه در نظر گرفته می شود، اگر درجه نسبت داده شده به یک نمونه، از سطح آستانه در نظر گرفته شده بیشتر باشد، به عنوان مثبت بودن نمونه تلقی می شود و بالعکس. در مورد این نوع طبقه بندی کننده ها در نمودار ROC با در نظر گرفتن سطوح مختلف آستانه چندین نقطه رسم می شود. اگر تعداد نمونه ها کم باشد این منحنی به صورت پله ای است. نمونه ای از نمودار ROC برای چنین طبقه بندی کننده هایی در شکل زیر مشاهده می شود.



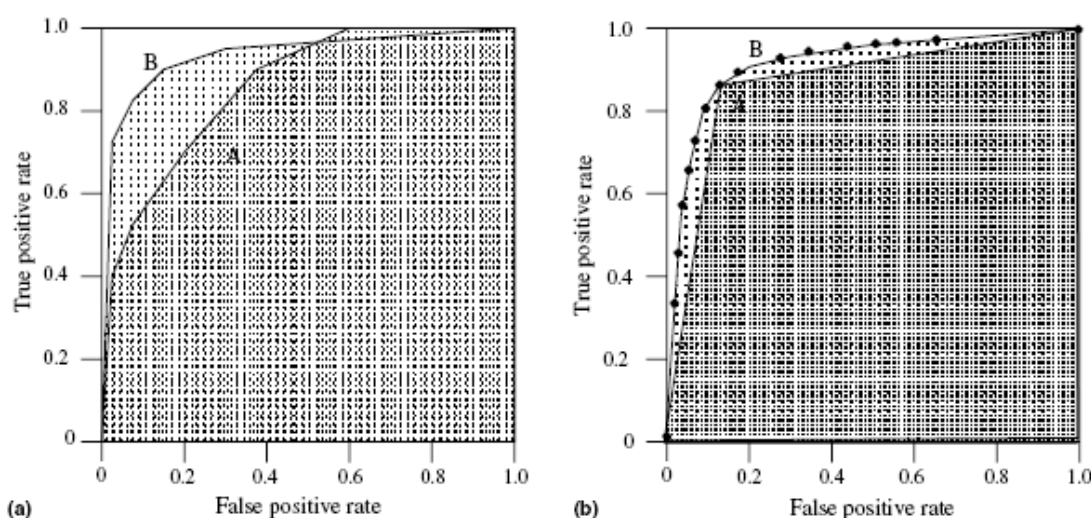
شکل ۱۲: نمودار ROC ایجاد شده به وسیله آستانه گذاری بر روی یک مجموعه داده تست. جدول سمت راست ۲۰ داده به همراه درجه های نسبت داده شده به هر کدام را نشان می دهد. نمودار نشان دهنده منحنی ROC متناظر است که هر نقطه با آستانه ای که تولید کننده ی آن نقطه است برچسب خورده است [۱۰].

نمودار ROC دارای خواص جالبی است، یکی از این خواص این است که این نمودار نسبت به تغییر در توزیع داده ها غیر حساس است. به این مفهوم که اگر نسبت نمونه های مثبت و منفی در داده های تست تغییر کند، منحنی ROC تغییری نخواهد کرد.

همانطوری که اشاره شد، طبقه بندی کننده هایی که خروجی گسسته (برچسب کلاس) تولید می کنند، در نمودار ROC فقط یک نقطه تولید می کنند که بر اساس مکان آن نقطه در نمودار ROC، قدرت عملکرد طبقه بندی کننده بررسی می شود. اگر در این نوع طبقه بندی کننده ها هدف تولید یک منحنی ROC به

جای یک نقطه باشد، باید طبقه بندی کننده را طوری تغییر داد که در خروجی به جای برچسب کلاس، یک درجه (احتمال) به هر ورودی نسبت دهد. این تغییر را با توجه به نحوه عملکرد هر طبقه بندی کننده می توان انجام داد.

منحنی ROC، یک ابزار نمایشی دو بعدی، برای ارزیابی یک طبقه بندی است. پارامتر AUC که همان سطح زیر منحنی ROC است، یک عدد اسکالر را برای مقایسه ی طبقه بندی کننده های مختلف ارائه می دهد. از آنجایی که AUC قسمتی از یک مربع واحد است، مقدار آن عددی بین ۰ تا ۱ است. همانطوری که در بخشهای قبلی دیده شد، زمانی که طبقه بندی به صورت تصادفی انجام شود، نقاط ایجاد شده در نمودار ROC، بر روی یک خط قطری بین (۰،۰) و (۱،۱) قرار می گیرند که سطح زیر آن ۰،۵ است. به این ترتیب یک طبقه بندی غیر قابل قبول باید دارای مقدار AUC کمتر از ۰،۵ باشد. AUC یک پارامتر مهم آماری به شمار می رود. شکل ۱۳ قسمت a سطح زیر دو منحنی ROC مربوط به دو طبقه بندی کننده را نشان می دهد. طبقه بندی کننده B سطح زیر منحنی بیشتری دارد، بنابراین متوسط نتایج بهتری نسبت به A خواهد داشت. البته گاهی ممکن است یک طبقه بندی کننده با مقدار AUC بالاتر، در یک نقطه بدتر از یک طبقه بندی کننده با AUC کمتر عمل کند. در شکل ۱۳ (a) این مسئله نشان داده شده است. طبقه بندی کننده B به طور کلی بهتر از A عمل می کند، به جز در $FPrate > 0.6$ که A یک برتری اندک بر B پیدا می کند. اما در عمل زمانی که یک تخمین کلی از قدرت پیشبینی مدل مد نظر باشد، AUC یک ابزار بسیار مناسب به شمار می آید. شکل ۱۳ قسمت b سطح زیر منحنی ROC برای دو طبقه بندی کننده با خروجی گسسته A و خروجی درجه بندی شده B را نشان می دهد. طبقه بندی کننده A همان طبقه بندی کننده B است فقط به ازای یک آستانه ثابت در نظر گرفته شده است. همانطوری که مشاهده می شود به علت رسم نشدن کامل نمودار ROC مقدار AUC در طبقه بندی کننده A کمی کمتر شده است. اگر از این مقدار صرف نظر کنیم، می توانیم مقدار AUC را به این ترتیب برای طبقه بندی کننده هایی که خروجی گسسته (برچسب کلاس) تولید می کنند بدست بیاوریم، بدون اینکه تغییری در ساختار طبقه بندی کننده به وجود بیاوریم.



شکل ۱۳: (a): سطح زیر دو نمودار ROC ایجاد شده به وسیله دو طبقه بندی کننده، (b): سطح زیر منحنی ROC یک طبقه بندی کننده ی گسسته A، و یک طبقه بندی کننده با خروجی درجه بندی شده B [۱۰].

همانطوری که اشاره شد، معمولاً منحنی ROC در مورد مسائلی با طبقه بندی دو کلاس به کار می رود و دو محور این منحنی نشان دهنده مصالحه بین خطا (تشخیص به اشتباه مثبت) و سود (تشخیص درست مثبت) است که طبقه بندی کننده بین دو کلاس ایجاد می کنند. در مورد مسائلی با بیشتر از دو کلاس رسم منحنی ROC کمی پیچیده می شود. با n کلاس جدول احتمال، یک ماتریس n در n خواهد شد که شامل n طبقه بندی صحیح است (قطر اصلی ماتریس) و $n^2 - n$ خطا وجود خواهد داشت (عناصر قطر غیر اصلی). در این موارد به جای مصالحه بین تشخیص درست TP و تشخیص به اشتباه مثبت FP، n سود و $n^2 - n$ خطا خواهیم داشت. با وجود سه کلاس یک ابرسطح $3^2 - 3 = 6$ به وجود می آید.

یک روش برای بررسی مسائل n کلاسه، تولید n نمودار ROC به ازای هر کلاس است. اگر C مجموعه ای از تمامی کلاس ها در نظر گرفته شود، نمودار i ام نمایش دهنده کیفیت یک طبقه بندی است، زمانی که کلاس c_i به عنوان کلاس مثبت و باقی کلاسها به عنوان کلاس منفی در نظر گرفته شود. در این مسائل برای بدست آوردن سطح زیر منحنی ROC فرمولهایی ارائه شده است که در اینجا به یک مورد از آنها اشاره می شود:

$$AUC_{total} = \frac{2}{|C|(|C|-1)} \sum_{(c_i, c_j) \in C} AUC(c_i, c_j)$$

در این رابطه $AUC(c_i, c_j)$ مقدار سطح زیر منحنی ROC است زمانی که فقط دو کلاس c_i و c_j در نظر گرفته شود. این فرمول نسبت به فرمولهای ارائه شده دیگر، نسبت به توزیع کلاسها حساس نمی باشد و به این ترتیب یک مقدار عددی برای سطح زیر ابر صفحه ای که مشاهده آن به آسانی ممکن نمی باشد را ارائه می دهد [۱۰].

* در محیط مطلب برای رسم این نمودار به ازای دو گروه از دستور **roc** و برای محاسبه ی سطح زیر این منحنی به ازای دو گروه از دستور **auroc** استفاده می شود. جعبه ابزار ROC از آدرس زیر قابل برداشت است:

<http://theoval.cmp.uea.ac.uk/~gcc/matlab/>

۶. تبدیل Z-Score

تعداد انحراف استاندارد هایی که یک داده از میانگین آن توزیع فاصله دارد، توسط پارامتری به نام z-score تعیین می شود. این پارامتر توسط رابطه زیر محاسبه می شود:

$$z = \frac{x - \mu}{\sigma}$$

در این رابطه μ میانگین توزیع و σ انحراف استاندارد توزیع و x یک داده از توزیع است که قرار است فاصله آن از میانگین بر حسب انحراف استاندارد محاسبه شود.

در مورد داده های نرمال زمانی که یک مقدار از z-score داده می شود، فاصله بین میانگین و z-score داده شده، یک نسبت ثابت و مشخص از سطح زیر منحنی توزیع داده ها را در بر میگیرد. برای مثال در یک توزیع نرمال، ۰,۳۴۱۳ ام از سطح زیر منحنی توزیع داده ها بین میانگین و $z=1$ قرار دارد. به این ترتیب در داده های نرمال به ازای هر مقدار z بدست آمده می توان با استفاده از جدول مقادیر z اظهار نمود که چه مقدار از سطح زیر منحنی داده ها بین میانگین و آن مقدار از z قرار دارند [۱۱].

علاوه بر این استفاده از تبدیل داده x به مقدار z ، در زمانهایی که به دنبال مقایسه مکان یک داده، نسبت به میانگین یک توزیع هستیم، نیز مفید می باشد. برای مثال اگر در یک مسابقه سنجش زیبایی یک فرد، نمره

۸۵ بگیرد، زمانی این نمره خوب خواهد بود که اکثر افراد نمره ۵۰ گرفته باشند؛ ولی اگر تقریباً بیشتر افراد نمره ۱۰۰ گرفته باشند، نمره ۸۵، نمی تواند بیانگر زیبایی فرد باشد. به بیان دیگر در موضوعات مختلف، معنای هر نمره و درجه ی بدست آمده در مقایسه با میانگین گروه معنا پیدا می کند. مسئله دیگر مربوط به زمانهایی است که بخواهیم درجه هایی که در واحدهای مختلف و یا از جامعه های متفاوت اندازه گیری شده اند را مورد مقایسه قرار دهیم. برای مثال چگونه می توان نمره ی ۸۵ مربوط به مسابقه سنجش زیبایی را با نمره ی ۱۰۰ در تست I.Q مورد مقایسه قرار داد. قبل از مقایسه درجات مربوط به توزیع هایی متفاوت (مانند مثال ذکر شده)، باید این درجات/ستاندارد شوند. برای منظور می توان از تبدیل داده ها به Z-score استفاده نمود. با توجه رابطه ی مربوط به محاسبه ی Z-score، زمانی که یک داده به Z-score تبدیل می شود، از میانگین کم و بر انحراف استاندارد جامعه مربوط به خود تقسیم می شود. بنابراین تمامی توزیع های Z-score دارای میانگین صفر و انحراف استاندارد یک هستند، به همین دلیل است که به کمک Z-score می توان دو داده از توزیع های مختلف را با هم مورد مقایسه قرار داد [۱۲].

به این ترتیب می توان اهداف محاسبه ی Z-score را به صورت زیر خلاصه نمود [۱۳]:

- **تعیین مکان یک داده از توزیع نسبت به میانگین.** همیشه Z-score دارای دو بخش است که

مکان یک داده را در توزیع مشخص می کند:

○ علامت Z-score: علامت Z-score مشخص می کند که یک داده در سمت راست (+) و یا در

سمت چپ (-) میانگین قرار دارد.

○ مقدار عددی Z-score: مقدار Z-score مشخص می کند که داده ی مورد نظر چند انحراف

استاندارد از میانگین فاصله دارد.

- **استاندارد کردن داده های گرفته شده از توزیع های متفاوت.** استاندارد کردن توزیع ها به ما

امکان مقایسه دو داده از توزیع های متفاوت را می دهد. شکل توزیع Z-score مشابه با شکل داده

هایی است که Z-score از آنها محاسبه شده است. میانگین توزیع Z-score همیشه صفر و انحراف

استاندارد آن یک است. به این ترتیب مقدار Z-score برابر با تعداد انحراف استاندارد هایی است که

از میانگین توزیع Z-score فاصله دارد.

* در محیط مطلب برای اجرای این تبدیل از دستور *zscore* استفاده می شود.

[با توجه به توانایی Z-score در رسیدن به اهداف ذکر شده، در مطالعات مربوط به پردازش سیگنالهای مغزی

از این پارامتر، به منظور های مختلفی استفاده شده است [۱۷-۱۴]. تیچر^۱ (۱۹۹۸) برای اولین بار از مفهوم

Z-score به عنوان بیوفیدبک EEG استفاده نمود و در ادامه این مطالعه، تحقیقات مختلفی بر روی این مسئله

صورت گرفت و امروزه از مقادیر Z-score به عنوان یک روش جدید در ایجاد بیوفیدبک و آموزش افراد به

خصوص در کاربردهای BCI استفاده می شود. در این موارد Zscore مربوط به ویژگی های انرژی مطلق و

نسبی، نسبت انرژی تتا به بتا، وابستگی بین نیمکره های مغزی، تقارن بین نیمکره ها و فرکانسهای مغزی،

محاسبه شده است. سپس از مقادیر Z-score بدست آمده جهت تعیین سطح افراد در حال آموزش، نسبت به

میانگین استفاده شده است (این میانگین از یک جامعه از افراد با شرایط نرمال محاسبه شده است).

همیشه وقتی منحنی فیت می کردیم

برای اینکه منحنی فیت شده به دو گروه رو با هم مقایسه کنیم می‌توانیم پارامترهای مربوط به منحنی های فیت شده رو مقایسه می کردیم

الان فهمیدیم می‌تونیم از بخش رگرسیون نرم افزارهای آماری استفاده کنیم و کلا نمودار ها را مقایسه آماری انجام بدیم

یه چیز دیگه که یاد گرفتیم اینه که تو ای ن نرم افزارها یه بخش دیگه هم هست به اسم کورلیشن

مثلا دو تا پارامتر داری می‌خوای ببینی کورلیشن بین اینها معنادره یا نه

بدون اینکه بینشون یک منحنی فیت کنی

اگه بینشون منحنی فیت کنی میشه رگرسیون

و باید از اماره های رگرسیون استفاده کنی

کورلیشن و رگرسیون به نظر میاد خیلی بهم شبیه هستن

ولی از نظر آماری تقریبا همیشه وقتی از کورلیشن استفاده می‌کنیم که هر دو تا متغیرمون رو اندازه گیری کرده باشیم

و نباید یکیش پارامتری باشه که در طی آزمایش خودمون هی تغییرش دادیم

ولی در رگرسیون معمولا یک پارامتری هست که خودمون هی تغییرش می‌دیم و با تغییرش یک متغیر رو اندازه می‌گیریم

Now

۷. مراجع

- [۱]. ح. توحیدی، ح. نظام آباد پور، س. سریزدی، "انتخاب ویژگی با استفاده از جمعیت مورچگان باینری"، اولین کنگره فازی و سیستم های هوشمند، دانشگاه فردوسی مشهد، ۱۳۸۶.
- [2]. W.Siedlecki and J.Sklansky, "On automatic feature selection", International Journal of Pattern Recognition and Artificial Intelligence, vol. 2, no. 2. pp. 197-220, 1989.
- [۳]. ر. آ. جانسون، د. د. ویچرن، ترجمه: ح. نیرومند، "تحلیل آماری چند متغیری کاربردی"، انتشارات دانشگاه فردوسی مشهد، صص. ۵۹۳-۷۹۴، ۱۳۷۸.
- [۴]. ا. فتوحی، ف. اصغری، "کتاب آموزشی آنالیز آماری داده ها در SPSS11"، تهران: کانون نشر علوم، صص. ۱۳۵-۴۳۰، ۱۳۸۲.
- [5]. Engineering Statistics Handbook, <http://www.itl.nist.gov/div898/handbook/eda/section3/eda33.htm>
- [6]. Jennifer Nord, "Box-And-Whisker Plot", 1995, <http://ellerbruch.nmu.edu/cs255/jnord/lineplot.html>
- [7]. "Box Plot", HyperStat Online Contents, http://www.statisticssolutions.com/landing.php?Source=David_Lane&Campaign=David_Lane_Banner&AdGroup=%20David_Lane_Banner%20&Keyword=%20David_Lane_Banner%20&Landing=index&Max_CPC=0&Type=None
- [۸]. م. هاشمی پرست، "آمار و احتمال در مهندسی و علوم"، انتشارات دانشگاه صنعتی خواجه نصیرالدین طوسی، صص. ۶۱۰-۶۱۸، ۱۳۸۵.
- [9]. Hogg, R. V., and J. Ledolter, Engineering Statistics. MacMillan, 1987.
- [10]. Tom Fawcett, "An introduction to ROC analysis", Elsevier, Pattern Recognition Letters **27** (2006) 861-874
- [11]. "What is a Z score?", <http://www.econ.iastate.edu/classes/crp274/swenson/CRP272/Davids%20Lectures/What%20is%20a%20Z%20score.pdf>
- [12]. H. Abdi, "Z-scores", Encyclopedia of Measurement and Statistics, 2007
- [13]. "About Z-Scores", Numeracy Project of University of Guelph, http://atrium.lib.uoguelph.ca/xmlui/bitstream/handle/10214/1842/A_About%20Z-Scores.pdf?sequence=2
- [14]. K. Nagata, "Topographic EEG Mapping in Cerebrovascular Disease", Brain Topography, Volume 2, Numbers 1/2, pp.119-128, 1989
- [15]. F. B. Vialatte, J. S. Casals, A. Cichocki, "EEG windowed statistical wavelet scoring for evaluation and discrimination of muscular artifacts", Physiol. Meas. **29** (2008) 1435-1452
- [16]. H.R. Jang, H.K. Ko, C. F. Vincent Latchoumane, J. H. Chae, J. Jeong, "Comparison of linear and nonlinear functional connectivity in Alcoholic patients", World Congress on Medical Physics and Biomedical Engineering 2006, IFMBE Proceedings, (2006) Vol.14, pp. 1115-1118
- [17]. R. W. Thatcher, C. J. Biver, D. M. North, "Z score EEG Biofeedback: Technical Foundations", Applied Neuroscience, Inc. (2007)