



Effect of normal distribution and least squares method on error theory

Mohammad Reza Majidi

Bachelor's degree student in mapping, Payame Noor University, North Tehran

Email:

majidi.mfa@gmail.com

Abstract

In the Least-Square Fitting method, instead of using the absolute magnitude of the error, we consider its square. Therefore, minimizing the total number of error squares leads to the magnification, which considers smaller errors. The most common way to determine the parameters that determine the curve is to determine the direction of error reduction and a small step in that direction and repeat the process until we reach convergence. This process of resolving parameters repeatedly is also known as the "gradient descent" method. In this tutorial, we use basic matrix calculations and use them to get the parameters for the best fit of the curve.

Keywords: Errors, Least Squares, Regression, Normal Distribution, Measurement



www.cpjournals.com

نشریه عمران و پروژه Civil & Project Journal(CPJ)

تأثیر توزیع نرمال و روش کمترین حداقل مربعات در تئوری خطاها

محمد رضا مجیدی

دانشجوی کارشناسی رشته نقشه برداری، دانشگاه پیام نور، تهران شمال

پست الکترونیکی:

majidi.mfa@gmail.com

تاریخ دریافت: ۱۳۹۹/۰۲/۳۱ تاریخ پذیرش: ۱۳۹۹/۰۳/۳۱

چکیده

برازش منحنی به برازش یک تابع از پیش تعریف شده اطلاق می‌شود که متغیرهای مستقل و وابسته را به یکدیگر مربوط می‌کند. گام اول در محاسبه بهترین منحنی یا خط، پارامتری کردن تابع خطا با استفاده از متغیرهای اسکالر کمتر، محاسبه مشتق خطا نسبت به پارامترها و در نهایت محاسبه پارامترهایی است که تابع هزینه خطا را کمینه می‌کنند. در روش «برازش حداقل مربعات (Least-Square Fitting)»، به جای آنکه از قدر مطلق خطا استفاده کنیم، مربع آن را در نظر می‌گیریم بنابراین، کمینه‌سازی مجموع مربعات خطا منجر به برازشی می‌شود که خطاهای کوچک‌تری را در نظر می‌گیرد. رایج‌ترین روش برای تعیین پارامترهایی که منحنی را مشخص می‌کنند، تعیین جهت کاهش خطا و یک گام کوچک در آن جهت و تکرار فرایند تا جایی است که به همگرایی برسیم. این فرایند حل تکراری پارامترها به عنوان روش «گرادیان کاهشی (Gradient Descent)» نیز شناخته می‌شود. در این آموزش، از محاسبات ماتریسی پایه استفاده می‌کنیم و آن‌ها را برای به دست آوردن پارامترها به منظور بهترین برازش منحنی به کار می‌گیریم.

کلمات کلیدی: خطاها، حداقل مربعات، رگرسیون، توزیع نرمال، اندازه گیری

۱- مقدمه

پایه و اساس همه ی کارهای نقشه برداری، اندازه گیری است؛ به ویژه اندازه گیری طول و زاویه که بیشتر آن در روی زمین و با وسایل نقشه برداری انجام می گیرد. بنابراین مجموعه ی عوامل انسان، محیط و دستگاه در نتایج اندازه گیری ها سهیم می باشند و هر کدام از این سه عامل، در به دست آمدن اندازه های دقیق و خوب یا برعکس، نتایج غیر دقیق و اشتباه، مؤثر هستند تمامی اندازه گیری های انجام گرفته اعم از اندازه گیری طول، زاویه و ارتفاع معمولاً همراه با یکسری خطاها همراه است. تشخیص و تعیین تمامی این خطاها مستلزم دانستن جنس و نوع آن می باشد. بطور کلی در مفهوم تئوری خطاها با دو نوع اصطلاح برخورد خواهیم داشت.

-اشتباه

-خطا

توزیع نرمال، که ممکن است، نمودار آن را به عنوان منحنی زنگدیس بشناسند، گاهی با نامهای پیرلاپلاس و کارس گاوس، که در تاریخ پیدایش آن نقش چشمگیر داشته اند، همراه است. گاوس توزیع نرمال را با روش ریاضی به عنوان توزیع احتمال خطای اندازه گیریها به دست آورد و آن را «قانون نرمال خطاها» نامید. بعداً منجمین، فیزیکدانها، و کمی بعد از آن، کسانی که در بسیاری از رشته ها داده ها را گردآوری می کردند، دریافتند که بافت نگارهای این داده ها دارای این خصوصیت مشترک هستند که ارتفاع مستطیلها ابتدا بتدریج به یک مقدار بیشینه صعود می کنند و سپس به طور متقارن کاهش می یابند. هرچه منحنی نرمال تنها منحنی نیست که چنین شکلی دارد ولی معلوم شده است که در موارد بسیار زیادی، تقریب قابل قبولی به دست می دهد. زمانی در جریان مراحل اولیه تکامل آمار، چنین احساس می شد که داده های مربوط به هر پدیده واقعی باید مطلقاً با منحنی نرمال زنگدیس باشند و در غیر این صورت می باید نسبت به فرایند جمع آوری داده ها مشکوک بود. از اینجاست که این توزیع به نام توزیع نرمال معروف شده است. لکن بررسی دقیق داده ها در اغلب موارد، نارسایی توزیع نرمال را آشکار ساخته است. لکن بررسی دقیق و در حقیقت، عمومیت توزیع نرمال افسانه ای بیش نیست، و مثالهای توزیع های غیرنرمال در هر یک از قلمروهای تحقیقات، فراوان اند. با وجود این، توزیع نرمال نقشی اساسی در آمار بازی می کند، و روشهای استنباطی که از آن به دست می آیند، دارای قلمرو کاربرد وسیعی هستند و ستون فقرات روشهای جاری تجزیه و تحلیل آماری را تشکیل می دهند. هرچند در اینجا صحبت از اهمیت توزیع نرمال است، ولی بحث ما در واقع به رده وسیعی از توزیعها که دارای چگالی زنگدیس اند،

مربوط می شود. هر توزیع نرمال به وسیله مقدار میانگین آن، و انحراف معیار آن، به طور کامل مشخص می شود؛ این مقادیر در فرمول تابع چگالی احتمال ظاهر می شوند.

۲- مروری بر مبانی نظری

رگرسیون^۱

در مدل های آماری، تحلیل رگرسیون یا تحلیل ارتباط یک فرایند آماری برای تخمین روابط بین متغیرها می باشد. این روش شامل تکنیک های زیادی برای مدل سازی و تحلیل متغیرهای خاص و منحصر بفرد، با تمرکز بر رابطه بین متغیر وابسته و یک یا چند متغیر مستقل، می باشد. تحلیل رگرسیون خصوصاً کمک می کند در فهم اینکه چگونه مقدار متغیر وابسته با تغییر هر کدام از متغیرهای مستقل و با ثابت بودن دیگر متغیرهای مستقل تغییر می کند. بیشترین کاربرد تحلیل رگرسیون تخمین امید ریاضی شرطی متغیر وابسته از متغیرهای مستقل معین است که معادل مقدار متوسط متغیر وابسته است وقتی که متغیرهای مستقل ثابت هستند. کمترین کاربرد آن تمرکز روی چندک یا پارامتر مکانی توزیع شرطی متغیر وابسته از متغیر مستقل معین است. در همه موارد هدف تخمین یک تابع از متغیرهای مستقل است که تابع رگرسیون نامیده شده است. در تحلیل رگرسیون تعیین پراکندگی متغیر وابسته اطراف تابع رگرسیون مورد توجه است که می تواند توسط یک توزیع احتمال توضیح داده شود.

توزیع نرمال

نام دیگر توزیع نرمال، «توزیع طبیعی» یا «تابع گاوسی» است، زیرا این تابع را نخستین بار کارل فردریش گاوس پیشنهاد کرده است. این توزیع یکی از مهمترین توزیع های احتمالی پیوسته در نظریه احتمالات است. علت نام گذاری و همچنین اهمیت این توزیع، همخوانی بسیاری از مقادیر حاصل شده، هنگام نوسان های طبیعی و فیزیکی پیرامون یک مقدار ثابت با مقادیر حاصل از این توزیع است. در ادامه انواع مختلفی از توزیع های محتمل داده ها را نشان داده ایم.

صحت

نزدیکی توافق بین مقدار میانگین حاصل از تعداد زیادی از نتایج آزمون و مقدار مرجع پذیرفته شده. به صحت «درستی میانگین» نیز گفته می شود ولی استفاده از آن به این صورت توصیه نمی شود. اندازه صحت معمولاً برحسب اربیبی بیان می شود.

^۱ regrasion

اریبی

اختلاف بین مقادیر موردانتظار نتایج آزمون و مقدار مرجع پذیرفته شده.

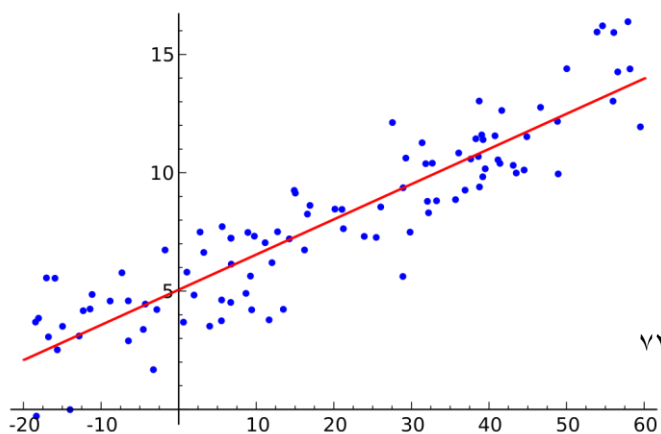
دقت

نزدیکی توافق بین نتایج حاصل از آزمون‌های مستقل تحت شرایط قراردادی. معمولاً اندازه دقت بر حسب عدم دقت بیان می‌گردد و بر حسب انحراف معیار نتایج آزمون محاسبه می‌شود. دقت پایین تر ناشی از انحراف معیار بالاتر است.

«نتایج آزمون مستقل» نتایجی است که تحت تاثیر هیچیک از نتایج آزمون قبلی که روی همان نمونه آزمون یا مشابه آن انجام شده است، قرار نمی‌گیرد. اندازه‌های کمی دقت وابستگی شدیدی به شرایط تصریح شده دارند. شرایط تکرارپذیری و تجدیدپذیری، دو حد نهایی دقت هستند.

تفاوت بین دقت و صحت: دقت فقط به توزیع خطاهای تصادفی بستگی دارد و ارتباطی به مقدار واقعی یا مقدار مشخص شده، ندارد. صحت که بر حسب اریبی بیان می‌شود. خطای سیستماتیک کل است که امکان دارد از یک یا چند مؤلفه خطای سیستماتیک تشکیل شده باشد. بزرگ بودن اریبی بیانگر اختلاف زیاد از مقدار مرجع پذیرفته شده است. صحت (اریبی) شامل خطای تصادفی نمی‌شود.

حداقل مربعات معمولی روش خط



از میان روشهای مختلف خطیبرای برآورد پارامترهای مدل (β) روش [۴] OLS یا همان حداقل مربعات معمولی به خاطر خواص مطلوبی (البته هنگامی که فروض این روش برقرار باشند) که دارد به عنوان پرکاربردترین و غالبترین روش شناخته شده است. این روش که به کارل فردریک گوس، ریاضیدان نامی آلمان نسبت داده می‌شود از طریق حداقل کردن مجموع مربعات جملات اخلال، تلاش میکند که بهترین خط رگرسیونی برای دادهها را برازش نماید..

$$y_i = \hat{\beta}_1 + \hat{\beta}_2 x_i + e_i$$

$$\min \sum e_i^2 = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)^2$$

$$\begin{cases} \hat{\beta}_1 = \frac{\sum x_i^2 \sum y_i^2 - \sum x_i \sum x_i y_i}{N \sum x_i^2 - (\sum x_i)^2} = \bar{y} - \hat{\beta}_2 \bar{x} \\ \hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2 - N \bar{x}^2} \end{cases}$$

اگر هدف ما تنها تخمین β_1 و β_2 باشد، روش OLS کافی است اما در تحلیل رگرسیون هدف تنها بدست آوردن برآوردهایی از β_1 و β_2 نیست بلکه هدف استنتاجاتی درباره β_1 و β_2 واقعی می‌باشد. برای رسیدن به این هدف نه تنها باید شکل تبعی مدل را تعیین کنیم بلکه باید فرضیات معینی درباره چگونگی بوجود آمدن u_i و x_i را نیز مطرح سازیم:

معادله (۱) نشان می‌دهد که y_i هم به u_i و هم به x_i بستگی دارد. بنابراین تا زمانی که ندانیم u_i و x_i چگونه بدست می‌آیند هیچ راهی برای دستیابی به استنتاجات آماری درباره y_i و همچنین β_1 و β_2 نخواهیم داشت. از این رو فرضیات مبتنی بر متغیرهای توضیحی و جزء خطا برای تفسیر معتبری از تخمینهای رگرسیون اهمیت دارند.

از آنجا که در روش OLS حداقل مربعات معمولی، پارامترهای تخمین زده شده به صورت ترکیب خطی از y_i هستند به همین خاطر روش OLS یک روش خطی نامیده میشود. برای نشان دادن اینکه پارامترهای تخمین زده شده ترکیب خطی از متغیر وابسته هستند، کافست متغیر α_i را به صورت زیر تعریف کنیم:

$$\alpha_i = \frac{x_i}{\sum x_i^2 - N \bar{x}^2}$$

در این صورت $\hat{\beta}_2$ برابر خواهد بود با:

$$\hat{\beta}_2 = \sum \alpha_i y_i$$

که ترکیب خطی از y_i می‌باشد.

روش OLS از مربعات جملات اخلال استفاده می‌کند. این بدین معناست که در روش OLS خطاهای بزرگتر بیشتر جریمه میشوند.

۳- چارچوب توزیع نرمال و روش کمترین حداقل مربعات در تئوری خطاها

منظور از اشتباه طبق تعریف انحرافی است که در نتیجه بی توجهی - فراموشی - بی تجربگی شخصی و به صورتهای مختلف ظاهر میگردد. در برخورد با کمیتی که احتمال وجود اشتباه در آن وجود دارد می بایستی آن کمیت از داخل تمامی مشاهدات صورت گرفته حذف گردد.

اما منظور از خطا به اختلاف میان کمیت اندازه گیری شده با مقدار واقعی یا مقدار میانگین را خطا می نامیم. در مجموع خطاها به دو دسته سیستماتیک (تدریجی) و اتفاقی تقسیم می شوند.

خطاهای سیستماتیک (تدریجی) به خطاهایی گفته می شوند که ماهیت برداری داشته و بصورت کاهنده یا افزایشنده ظاهر می گردد. خطاهای اتفاقی به خطاهایی گفته می شوند که ماهیت عددی یا نرده ای دارند و امکان رفع و سرشکنی برای آن وجود خواهد داشت.

عمده ترین خطاهای تدریجی (سیستماتیک) به حالتهای ذیل طبقه بندی می گردند:

-خطاهای دستگاهی

-خطاهای طبیعی

-خطاهای انسانی

خطاهای دستگاهی اغلب به دلیل بدکار کردن و معیوب بودن وسایل اندازه گیری به وجود می آید و راه پیشگیری از آن ها کنترل مستمر وسایل می باشد. به این ترتیب که قبل از آن که وسایل را به منطقه ی موردنظر حمل کنیم باید در محل انبار سالم بودن و صحیح کار کردن آن ها را کنترل نماییم.

خطاهای طبیعی در نتیجه وضعیت آب و هوایی و شرایط اقلیمی منطقه جغرافیایی بوجود می آید.

خطاهای انسانی انسان در صورت عدم سلامت حواس پنج گانه ی خویش در انجام فعالیت های دقیق نقشه برداری همانند خواندن، نوشتن، محاسبه و ترسیم دچار اشتباه می گردد.

با همه ی تمهیدات و مواظبت هایی که انجام آن ها توصیه گردید باز هم امکان اشتباه در اندازه گیری ها وجود دارد؛ بنابراین، باید روش هایی به کار ببریم تا اگر در مواردی موفق نشده ایم از اشتباه در اندازه گیری ها جلوگیری نماییم، قبل از ترک محل عملیات صحرائی، اشتباهات موجود را پیدا کرده، در صورت نیاز، اندازه گیری های صحیح را به جای اندازه گیری های اشتباه به کار ببریم توجه کنید که در بعضی موارد، محل اندازه گیری ها ممکن است در نقطه ی بسیار دوری قرار گرفته باشد و هزینه مراجعه مجدد به محل زیاد باشد یا ممکن است کار از نظر زمانی به تأخیر بیافتد.

کنترل، بهترین راه برای مبارزه با اشتباه در نقشه برداری است، به همین دلیل در نقشه برداری، کار بدون کنترل فاقد ارزش است و استفاده از اندازه های کنترل نشده بسیار مضر و خطرناک می باشد.

کنترل، در کلیه ی مراحل تهیه ی نقشه، از شناسایی و اندازه گیری گرفته تا محاسبه و ترسیم اعمال می شود تا نتیجه کار نقشه برداری که همان نقشه است، دقیق و مطمئن باشد. در قسمت قبل، از کنترل تکنیکی عوامل انسانی و کنترل فنی وسایل صحبت کردیم. در این قسمت از کنترل اندازه های به دست آمده صحبت می کنیم. در نقشه برداری معمولاً نتایج اندازه گیری ها را به دو روش کنترل می کنیم.

کنترل از طریق تکرار اندازه گیری ها

در این روش برای کنترل اشتباه نبودن اندازه ی به دست آمده، اندازه گیری را دو یا چند بار انجام می دهند؛ مثلاً برای ب هدست آوردن اندازه ی یک زاویه دو یا چند بار آن را اندازه گیری می کنند یا برای اندازه گیری یک طول آن را به صورت رفت و برگشت اندازه گیری می نمایند.

کنترل از طریق مقایسه با یک مدل ریاضی

در این روش اندازه های به دست آمده را با یک مدل ریاضی مقایسه می کنیم؛ را به دست C و B و A اندازه ی سه زاویه ی ABC مثلاً اگر در مثلث آورده ایم می توانیم آن ها را با یک مدل ریاضی که براساس آن: 180° است محک بزنیم. مجموع سه

زاویه ی داخلی هر مثلث اگر مجموع سه زاویه ۱۸۰ درجه یا حتی خیلی نزدیک به آن بود احتمالاً اندازه های به دست آمده ممکن است بدون اشتباه باشند.

حقیقت این است که هم روش تکرار و هم روش مقایسه با مدل ریاضی به تنها یی کامل نیستند و معمولاً برای کارهای دقیق هر دو روش را با هم به کار می برند؛ یعنی مشاهده را چندبار اندازه گیری می کنند و انداز های را که با بقیه اختلاف زیادی دارد اشتباه تلقی کرده، حذف می کنند و هم این که اندازه های به دست آمده را با یک مدل ریاضی مقایسه می کنند تا اشتباهات پنهان را آشکار کنند.

توجه به این نکته ضروری است که هر چند اشتباه ممکن است آن قدر بزرگ باشد که حتی با یک بار اندازه گیری نیز قابل تشخیص باشد) مانند وقتی که اندازه عرض یک اتاق را که ظاهراً ۴ متر است ۴۰ متر به دست آورده باشند (اما غالباً پیدا کردن اشتباه به این سادگی امکان پذیر نیست و باید نقشه برداران با روش های کنترلی مناسب آن ها را یافته و حذف نمایند.

کشیدگی یا برجستگی

توصیف کننده ی میزان قله ای بودن یک توزیع احتمالی است. Kurtosis در آمار و نظریه احتمالات برجستگی، کشیدگی یا کشیدگی برابر با گشتاور چهارم نرمال شده است، به عبارت دیگر کشیدگی معیاری از تیزی منحنی در نقطه ماکزیمم است (حسنی پاک، ۱۳۸۶). مقدار کشیدگی برای توزیع نرمال برابر ۳ می باشد (جانسون و همکاران، ۲۰۰۱). کشیدگی بیان کننده نحوه انباشته شدن نمره ها در مرکز توزیع یک متغیر است. در واقع کشیدگی، برآمدگی یا فرورفتگی منحنی توزیع نرمال است. در حالت کشیدگی تقارن توزیع حفظ می شود و دو نیمه منحنی متقارن هستند، اما نقطه اوج منحنی نرمال دچار تغییر می شود. مقدار کشیدگی می تواند مثبت یا منفی باشد. در حالت کشیدگی منفی، منحنی توزیع متغیر فررفته تر از حالت نرمال می شود و نقطه اوج توزیع متغیر، پایین تر از توزیع نرمال است. در حالت کشیدگی مثبت، منحنی توزیع متغیر برآمده تر از منحنی نرمال می شود و نقطه اوج توزیع متغیر، بالاتر از توزیع نرمال است.

کجی یا چولگی

کجی یا چولگی در آمار، مقداری است از تقارن توزیع یک متغیر در اطراف میانگین. در واقع کجی، انحراف منحنی نرمال از حالت تقارن است. مقدار کجی می تواند مثبت یا منفی باشد. در حالت کجی مثبت، میانگین بزرگتر از میانه و میانه بزرگتر از مد است و در حالت کجی منفی، مد بزرگتر از میانه و میانه بزرگتر از میانگین است. در آمار و نظریه احتمالات چولگی نشان دهنده میزان عدم تقارن توزیع احتمالی است. اگر داده‌ها نسبت به میانگین متقارن باشند، چولگی برابر صفر خواهد بود. چولگی برابر با گشتاور سوم نرمال شده است. چولگی در حقیقت معیاری از وجود یا عدم تقارن تابع توزیع می باشد. برای یک توزیع کاملاً متقارن چولگی صفر و برای یک توزیع نامتقارن با کشیدگی به سمت مقادیر بالاتر چولگی مثبت و برای توزیع نامتقارن با کشیدگی به سمت مقادیر کوچکتر مقدار چولگی منفی است (جانسون و همکاران، ۲۰۰۱).

اما باید بدانیم که یک هیستوگرام یا منحنی فراوانی چه میزان به چپ یا به راست چوله است. برای اندازه گیری میزان چولگی می توان از روش زیر استفاده کرد:

ضریب چولگی گشتاوری: که در آن گشتاور مرتبه سوم حول میانگین نام دارد.

۱- اگر ضریب چولگی منفی باشد منحنی چوله به چپ است.

۲- اگر ضریب چولگی مثبت باشد منحنی چوله به چپ است.

۳- اگر قدر مطلق ضریب چولگی کوچکتر از ۰.۱ باشد توزیع داده ها نرمال است.

۴- اگر قدر مطلق ضریب چولگی بزرگتر از ۰.۱ و کوچکتر از ۰.۵ باشد توزیع داده ها تقریباً نرمال است.

۵- اگر قدر مطلق ضریب چولگی بزرگتر از ۰.۵ باشد توزیع داده ها نرمال نیست.

برای محاسبه ضریب چولگی یا کجی، فرمولهای زیادی وجود دارد. بعضی از آنها بر اساس داده ها و فراوانی آنها و با توجه به میانگین محاسبه می شود که عمدتاً همین روش کاربرد دارد. برخی دیگر، بر مبنای میانه یا نما می تواند محاسبه گردد.

۴- نتیجه گیری

چون توزیع نرمال، پیوسته، حول میانگین خود متقارن و زنگوله مانند است، به همین دلیل بسیاری از داده‌هایی که از پدیده‌های طبیعی به دست می‌آیند تقریباً (این تقریباً کمی تا قسمتی مهم است) از توزیع نرمال پیروی می‌کنند. برای مثال قد یک جامعه را در نظر بگیرید، در یک جامعه معمولاً تعداد کمی از افراد قد بسیار بلند و یا بسیار کوتاه دارند، و اکثریت جامعه قدی متوسط و نزدیک به میانگین دارند و نمودار فراوانی داده‌ها در آن شبیه زنگوله خواهد بود. (اگر به داده‌ای همانند قد یک افراد جامعه، وزن یا IQ دسترسی دارید می‌توانید این موضوع را بررسی کنید)، به همین دلیل می‌توان با استفاده از توزیع نرمال و خواص آن این پدیده‌ها را مورد تحلیل آماری قرار داد. دلیل دوم اهمیت توزیع نرمال قضیه حد مرکزی یا CLT است. خلاصه این قضیه این است که تحت شرایط خاص، میانگین (یا جمع) چند متغیر تصادفی که هر کدام واریانس و میانگین تعریف شده و محدود دارند، به طور تقریبی از توزیع نرمال پیروی می‌کند. این قضیه که باعث می‌شود بسیاری از محاسبات آماری تا حدی ساده‌تر شود، پایه‌ی بسیاری از محاسبات و روابط مربوط به نمونه‌گیری و کنترل کیفیت آماری است. پیش فرض‌های آماری، پایه بسیاری از آزمون - های آماری تک متغیری و چندمتغیری است. شرایط مهم واساسی برای تحلیل داده‌های چندمتغیری، برقراری پیش‌فرض‌های نرمال بودن، خطی بودن و یکسانی پراکندگی داده هاست. چنانچه یک یا چندتا از این مفروضه ها نادیده گرفته شود، در این صورت، در نتایج آماری سوگیری یا تحریف رخ می‌دهد (میزر و همکاران، ۱۳۹۱: ۱۰۷). قبل از انجام تحلیل‌های آماری تک متغیره و چندمتغیره (که پیش فرض‌های نرمال بودن، خطی بودن، یکسانی پراکندگی و نبود هم خطی چندگانه بین متغیرهای مستقل در مورد آن‌ها باید صدق کند) باید برقراری پیش فرض‌های آماری را بیازماییم. اگر انحراف از پیش فرض‌های آماری ناچیز باشد می‌توان با کمی تسامح و تساهل این انحراف را نادیده گرفت و به ادامه تحلیل پرداخت. اگر انحراف از پیش فرض‌ها قابل توجه باشد باید یا از روش تبدیل داده‌ها برای برقرار کردن مجدد پیش فرض‌ها استفاده کنیم یا از آزمون‌های جایگزین استفاده کنیم.

منابع

- جزوه تئوری خطا مهندس ابراهیم راستگو-۱۳۹۴
 جزوه تئوری خطا مهندس مهدی کریمی نژاد-۱۳۹۲
 کتاب تئوری خطا ها-ابولفضل رنجبر-۱۳۸۸
 کتاب هنرستان درس نقشه برداری-۱۳۸۹
 ماهنامه علمی تخصصی نقشه برداری - شماره ۳۲۵

- «An Etymological Dictionary of Astronomy and Astrophysics - ۱». dictionary.obspm.fr. -۰۶-۲۰۲۰
- «An Etymological Dictionary of Astronomy and Astrophysics - ۱». dictionary.obspm.fr. ۰۵-۰۶-۲۰۲۰.
- Fawcett, Tom (۲۰۰۶). "An Introduction to ROC Analysis" (PDF). Pattern Recognition Letters. ۲۷ (۸): ۸۶۱-۸۷۴. doi:۱۰.۱۰۱۶/j.patrec.۲۰۰۵.۱۰.۰۱۰.
- Powers, David M W (۲۰۱۱). "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation" (PDF). Journal of Machine Learning Technologies. ۲ (۱): ۳۷-۶۳.
- Rencher, Alvin C.; Christensen, William F. (۲۰۱۲-۰۸-۱۵). *Methods of Multivariate Analysis English*. John Wiley & Sons. p. ۱۹. ISBN ۹۷۸۱۱۱۸۳۹۱۶۷۹.
- Rencher, Alvin C.; Christensen, William F. (۲۰۱۲-۰۸-۱۵). *Methods of Multivariate Analysis*
- Yan, Xin (۲۰۰۹). *Linear Regression Analysis: Theory and Computing*. World Scientific. ISBN ۹۷۸۹۸۱۲۸۳۴۱۱۹.