تعیین صحت و دقت داده های آماری

فهرست مطالب **صفحه**

1- تعیین صحت و دقت داده ها.....2

شکل1-نمایش تغییرات صحت و دقت اندازه گیریها.....4

1-1- کنترل کیفیت نتایج آزمایشگاهی.....5

1-1-1- روش های گرافیکی.....6

شکل2- نمودار کنترل عملیات آزمایشگاهی.....7

جدول1- مقادیر اندازه گیری شده در جفت نمونه های تکراری در یک آزمایشگاه.....8

1-1-2- روش محاسباتی.....8

1-1-3- مقایسه میانگین دو سری اندازه گیری با استفاده از آزمون استیودنت-t.....9

1-1-4- مقایسه دو سری اندازه گیری با استفاده از تجزیه و تحلیل رگرسیون.....9

1- تعیین صحت و دقت داده ها

به منظور استفاده ی مطلوب از نتایج اندازه گیری ها(مثلا عیارسنجی) باید از درجه اعتمادی که می توان به آنها داشت آگاه بود. بدین منظور آزمایشات کنترل کیفیت جهت تعیین خطای مراحل آماده سازی و اندازه گیری انجام می شود. به طور کلی قابلیت اعتماد به داده های حاصل از اندازه گیری تابع مقدار دو نوع خطای تصادفی و سیستماتیک است.

خطاهای تصادفی در اثر عوامل خارج از کنترل شخص عمل کننده حاصل می گردند و بدین لحاظ وجود آنها (نه بزرگی آنها) اجتناب ناپذیر است. ولی با به کارگیری دستگاههای دقیق تر یا روش های مناسب تر می توان مقدار آن را کاهش داد. از ویژگی های آماری این نوع خطا این است که جمع جبری آنها حول میانگین باید صفر شود.

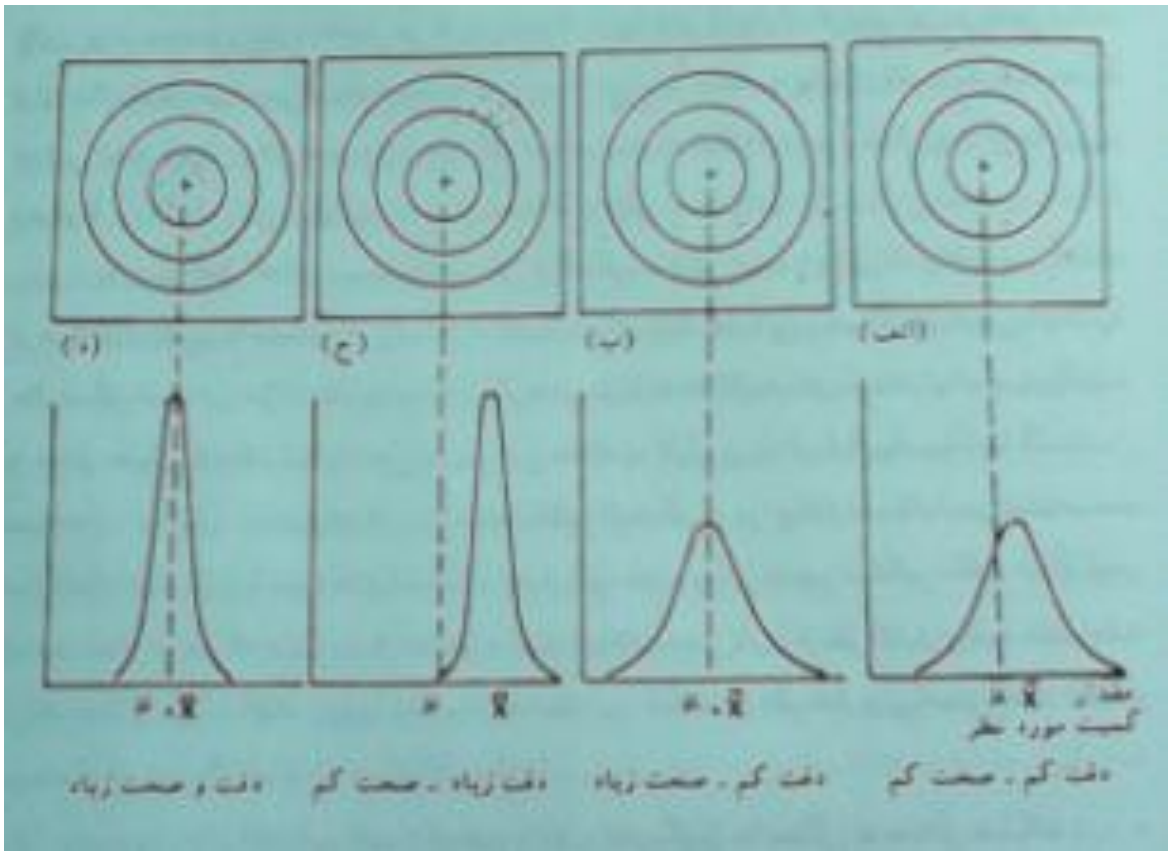
خطای سیستماتیک باعث می شود که میانگین مقادیر اندازه گیری شده یک کمیت به اندازه ای معین از مقدار حقیقی آن کمتر یا بیشتر شود. مقدار این خطا معمولا از قانون خاصی تبعیت می کند، ولی بدیهی است که جمع جبری آنها حول میانگین صفر نمی شود و این خود عامل اصلی انحراف مقدار میانگین از مقدار حقیقی است، زیرا خطای تصادفی صرفا در اثر تغییرات تصادفی ذاتی ایجاد می شود و در هر نوع اندازه گیری کم و بیش وجود دارد.

اگر مقدار یک کمیت معین، فقط با یک مرتبه آزمایش(یا نمونه برداری) تعیین شود، نمی توان گفت که اختلاف مقدار اندازه گیری شده با مقدار حقیقی آن در اثر ورود کدامیک از دو نوع خطای تصادفی یا سیستماتیک بوده است. اما اگر اندازه گیری ها به دفعات زیادی تکرار گردد، با محاسبه ی مقدار میانگین و پراش آنها می توان به بزرگی خطاهای تصادفی پی برد. در همین حال اختلاف بین مقدار میانگین اندازه گیری ها و مقدار حقیقی می تواند بزرگی اثر خطای سیستماتیک را نشان دهد. روش دیگری برای ارزیابی بزرگی این نوع خطا محاسبه انحراف معیار مقدار میانگین یک سری اندازه گیری است.

صحت یک اندازه گیری معرف نزدیکی مقدار اندازه گیری شده به اندازه ی حقیقی آن است. در واقع صحت یک اندازه گیری درجه درستی و صدق مقدار اندازه گیری شده را نشان می دهد ولی دقت بیان کننده ی این است که اندازه گیری های مختلف تا چه حد به هم نزدیک بوده اند. پس دقت را می توان معرف قابلیت تکرار با نتایج مشابه تلقی کرد. اگر مقدار واقعی یک کمیت را مشابه نقطه هدف در تیراندازی در نظر بگیریم و نتیجه هر اندازه گیری را مشابه اثر اصابت تیر روی صفحه هدف فرض کنیم آنگاه، به کمک شکل 1 می توان درک بهتری از مفهوم صحت و دقت و ارتباط آنها با خطاهای تصادفی و سیستماتیک بدست آورد.

در شکل(1 الف)، نقاط بدست آمده دور از هم هستند و میانگین آنها نیز با مقدار حقیقی کمیت مورد نظر(هدف) یعنی μ نیز فاصله زیادی دارد. بنابراین صحت عملیات کم است و این در حالی است که نقاط نیز فاصله زیادی از هم دارند پس دقت عملیات هم کم است. در شکل(1ب) نقاط پراکنده هستند ولی نحوه ی پراکندگی آنها حول مقدار حقیقی تقریبا یکنواخت است. در این حالت به علت فاصله زیاد مقادیر بدست آمده دقت عملیات کم، ولی چون میانگین این مقادیر منطبق با مقدار واقعی است، صحت عملیات زیاد است. البته باید توجه داشت که در عمل می توان فاصله x میانگین و μ را کم کرد ولی هرگز به طور مطلق نمی توان آنها را بر هم منطبق کرد.

در شکل (1ج) نقاط نزدیک هم هستند یعنی قابلیت تکرار با نتایج مشابه بهتر از حالت های قبل است ولی میانگین آنها با مقدار حقیقی کمیت فاصله زیادی دارد، پس دقت عملیات زیاد ولی صحت آن کم است. در عوض در شکل (1د) که مطلوب ترین حالت است، نقاط علاوه براین که به یکدیگر نزدیک هستند، به مقدار حقیقی هم نزدیک می باشند. پس دقت و صحت اندازه گیری ها زیاد است. همانطور که دیده می شود در حالت هایی که دقت عملیات کم است، منحنی توزیع حالتی بازتر و گسترده تر دارد. در صورتیکه هر چه اعداد حاصل از اندازه گیری به هم نزدیک تر باشند و در نتیجه دقت بیشتر باشد، این منحنی بسته تر خواهد شد.



شکل1: نمایش تغییرات صحت و دقت اندازه گیریها

خطای سیستماتیک می تواند باعث جابجایی منحنی و در نتیجه مقدار میانگین یک سری از اندازه گیریها(X میانگین) گردد. هر چه این جابجایی و در نتیجه انحراف از مقدار واقعی کمیت بیشتر باشد، مقدار خطای سیستماتیک بیشتر است. اگر اندازه گیریها در چند سری صورت گیرد و برای هر سری مقدار میانگینی محاسبه شود و سپس انحراف معیار مقدار میانگین محاسبه گردد می توان مقدار کمیت اخیر(انحراف معیار مقدار میانگین) را معرف صحت دانست. به طوری که هر چه این کمیت کوچکتر باشد صحت بالاتر است. ممکن است انحراف معیار مقدار میانگین را از روش تئوری برآورد نمود و از آن برای ارزیابی صحت اندازه گیریها استفاده کرد.

در پروژه های مختلف براساس درجه اهمیت اندازه گیریها حد مجاز صحت و دقت از پیش انتخاب می شود و طراحی ها براساس آنها صورت می گیرد. واضح است که هر چه بدنبال دستیابی به صحت و دقت بیشتری باشیم، باید هزینه زیاد تری را بپردازیم. بنابراین در هر پروژه ای با توجه به سقف هزینه های لازم برای بالا بردن صحت و دقت و ریسک های ناشی از پائین بودن دقت و صحت اندازه گیریها باید حالت بهینه ای را به عنوان حد مجاز

انتخاب کنیم. گاهی ممکن است زیانی که از پائین بودن صحت و دقت حاصل می شود بیشتر از هزینه های لازم برای افزایش این دو عامل باشد. برای مثال تعیین عیار مس در یک محموله ی 100000 تنی باید با صحت بیشتری نسبت به یک محموله ی ده تنی انجام شود. زیرا در صورت وجود یک خطای سیستماتیک ثابت، مقدار زیان احتمالی در محموله ی 100000 تنی به مراتب بیشتر از محموله ی 10 تنی است.

در عمل خطر خطای سیستماتیک بیش از خطای تصادفی است. زیرا اثر آن به طور نسبتا مشابه در همه ی اندازه گیریها منعکس می شود و امکان دارد برای مدت طولانی وجود آن تشخیص داده نشود. حال ممکن است این سوال مطرح شود که چگونه می توان حداقل خطای سیستماتیک دست یافت و به مقدار حقیقی نزدیک شد. راه حل تقریبی این مسئله به کارگیری روش کالیبراسیون با استفاده از نمونه های استاندارد است. برای بررسی صحت نتایج اندازه گیری هر چند وقت یکبار می توان صحت دستگاه اندازه گیری را با نمونه های استاندارد کنترل نمود. در صورتیکه بین میانگین مقادیر اندازه گیری شده و مقدار حقیقی که همان مقدار استاندارد است اختلاف معنی داری از نظر آماری وجود داشته باشد با یک خطای سیستماتیک روبرو هستیم و چنانچه این اختلاف از نظر آماری بی معنی باشد خطای سیستماتیک اندازه گیریها قابل اغماض است.

در پروژه های اکتشافی اهمیت صحت و دقت اندازه گیریها بستگی به مقیاس عملیات دارد. به عنوان مثال در برداشت های اکتشافی تا قبل از مرحله تعیین ذخیره آنچه که بیشتر اهمیت دارد دقت اندازه گیریها است، زیرا در این مقیاس از عملیات سنجش ها نسبی است نه مطلق و کاهش یا افزایش مقدار ثابتی به همه ی داده ها در وضعیت نسبی آنها بی تاثیر است. در مرحله ی تخمین ذخیره گرچه دقت و صحت هر دو مهم است ولی خطر اصلی کاهش صحت است.

1-1- کنترل کیفیت نتایج آزمایشگاهی

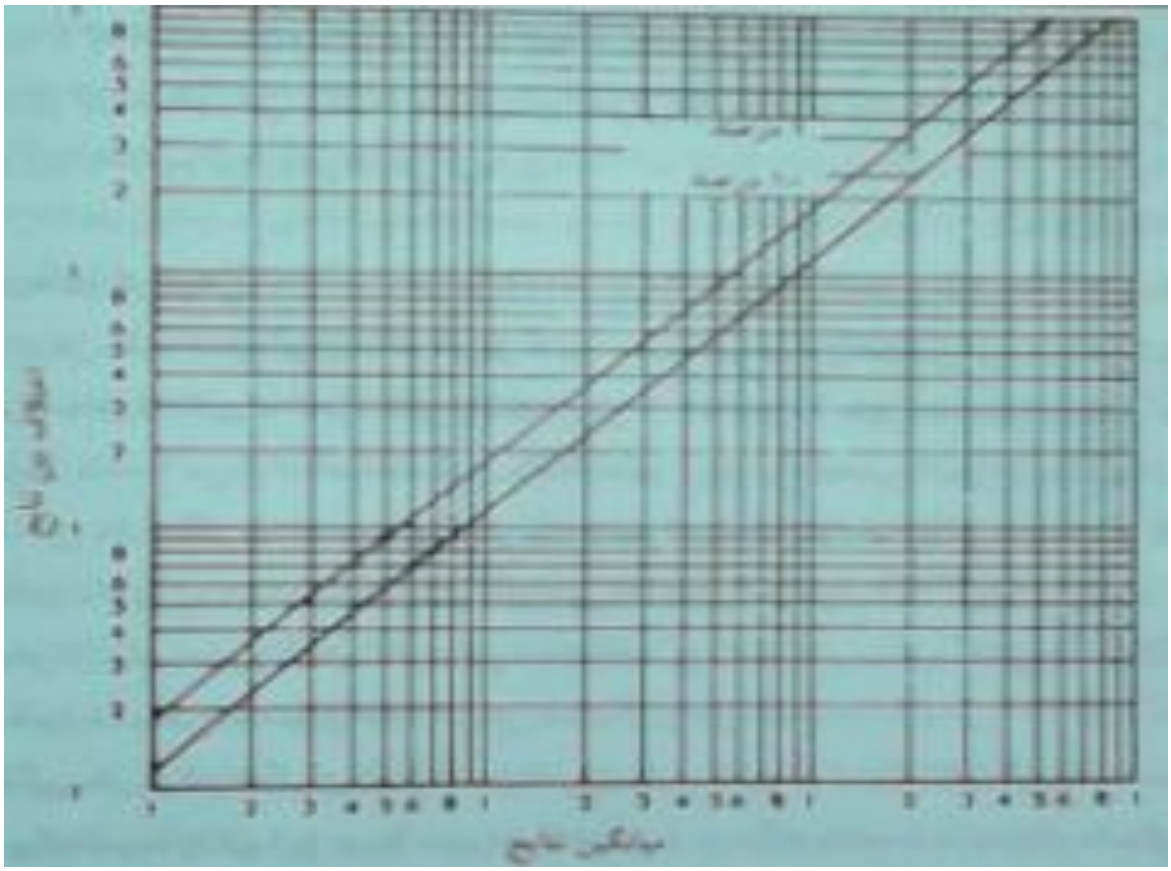
کنترل کیفیت داده های آزمایشگاهی از اهمیت ویژه ای برخوردار است. زیرا اولا میزان اعتماد به داده ها را مشخص می کند و ثانیا اگر خطای داده ها زیاد باشد(و نتوان اندازه گیریها را تکرار کرد) بهتر است در تفسیر نتایج دقت بیشتری به عمل آورد. برای تعیین دقت آزمایشگاه می توان یک سری نمونه ی تکراری تهیه کرد و به همراه نمونه

های اصلی به آزمایشگاه فرستاد و سپس دقت اندازه گیریها را محاسبه کرد. برای این منظور می توان از نمودار های کنترلی یا روابط محاسباتی مانند ستیودنت-t استفاده کرد، در ادامه به شرح این روش ها می پردازیم.

1-1-1- روش های گرافیکی

در این روش دقت اندازه گیریها از طریق آنالیز جفت نمونه های تکراری بررسی می شود. اگر تعداد جفت نمونه های تکراری به اندازه کافی باشد (بیش از 50 نمونه) تغییرات انحراف معیار در دامنه ی مقادیر اندازه گیری شده را می توان محاسبه کرد. دامنه ی مجاز برای این تخمین کمتر از ده درصد مقادیر اندازه گیری شده می باشد. اگر تعداد نمونه ها کمتر از 50 نمونه باشد (که معمولا در اکتشافات ناحیه ای حداکثر نمونه های تکراری 30 عدد است) باید از روشی که تامسون و هوارث بیان نموده اند استفاده کرد. در این روش از یک نمودار تمام لگاریتمی استفاده می شود که محور افقی آن میانگین دو اندازه گیری و محور قائم آن قدر مطلق اختلاف دو اندازه گیری می باشد.

اگر میانگین دو بار اندازه گیری روی محور افقی و قدر مطلق دو مقدار اندازه گیری شده در روی محور عمودی پیاده شود، تلاقی این دو به صورت نقطه ای در دستگاه مختصات نمایش داده می شود. معمولا در دستگاه مختصات مورد نظر دو خط مایل یکی برای 10 درصد و دیگری برای 1 درصد خطا از قبل رسم شده است. شکل 2 نمونه ای از این نمودار را نشان می دهد. اگر توزیع نقاط رسم شده در این دستگاه مختصات طوری باشد که 90 درصد آنها زیر خط 10 درصد 99 درصد آنها زیر خط 1 درصد قرار گیرند، خطای اندازه گیریها (به عنوان دقت) برابر 10 درصد برآورد می گردد.



شکل 2: نمودار کنترل عملیات آزمایشگاهی

جدول1: مقادیر اندازه گیری شده در جفت نمونه های تکراری در یک آزمایشگاه

$(Bi_1-Bi_2)\times 100$	$\frac{Bi_1+Bi_2}{2}\times 100$	Bi_2(ppm)	Bi_1 (ppm)	ردیف
۱۲	۲۹	۰/۲۳	۰/۳۵	۱
۲۰	۳۰	۰/۱۵	۰/۳۵	۲
۴	۲۸	۰/۲۴	۰/۴	۳
۲۴	۳۲	۰/۱۹	۰/۳۵	۴
۸۲	۸۱	۰/۴	۱/۲۲	۵
۱۶	۳۲	۰/۲۲	۰/۴	۶
۴	۲۸	۰/۲۸	۰/۴	۷
۶	۱۲	۰/۲۲	۰/۳	۸
۲۰	۲۵	۰/۱۵	۰/۳۵	۹
۲۱	۳۲.۵	۰/۲۲	۰/۳۵	۱۰
۱۱	۲۷.۵	۰/۲۲	۰/۳۳	۱۱
۱۶	۲۴.۵	۰/۲۵	۰/۴۲	۱۲
۲	۳۲	۰/۲۲	۰/۲۵	۱۳
۲۰	۳۰	۰/۴	۰/۴	۱۴
۲۲	۲۹	۰/۱۸	۰/۴	۱۵
۱۸	۳۶	۰/۲۶	۰/۳۵	۱۶
۱۹	۲۵.۵	۰/۱۶	۰/۳۵	۱۷
۱۶	۲۶.۵	۰/۱۸	۰/۳۵	۱۸
۱۶	۳۲	۰/۲۲	۰/۴	۱۹
۱۱	۳۶.۵	۰/۲۹	۰/۴	۲۰

2-1-1- روش محاسباتی

علاوه بر روش فوق یک روش ساده برای محاسبه ی خطای نسبی اندازه گیریها وجود دارد که با استفاده از رابطه ی زیر محاسبه می شود، این روش بیشتر برای محاسبه ی دقت عیارسنجی مناسبت دارد تا در اکتشافات ناحیه ای:

$$e = \frac{2}{n}\sum_{i=1}^{n}\frac{|x_{i-}y_i|}{x_{i+}y_i}$$

که در آن n تعداد نمونه های تکراری، x_i و y_i مقادیر اندازه گیری شده در نمونه های تکراری متناظر می باشند.

1-1-3- مقایسه میانگین دو سری اندازه گیری با استفاده از آزمون ستیودنت-t

در هر اندازه گیری خطای تصادفی وجود دارد. سعی می شود مقدار آن حتی الامکان تا حد مجاز پائین آید ولی به صفر نمی رسد. آنچه غیر قابل قبول است وجود خطاهای سیستماتیک در اندازه گیری است. لذا لازم است این نوع خطا شناسایی و حذف شود. فرض کنید نتایج حاصل از اندازه گیری تعدادی نمونه تکراری در دو آزمایشگاه مختلف در دسترس باشد. سوالی که مطرح می شود این است که آیا اختلاف بین این دو سری اندازه گیری (به عنوان نماینده دو جامعه ی مورد اندازه گیری در دو آزمایشگاه) معنی دار است یا خیر. به عبارت دیگر آیا اختلاف بین میانگین دو سری اندازه گیری تصادفی است یا نظام دار است. جواب این سوال با استفاده از آزمون ستیودنت-t داده می شود.

1-1-4- مقایسه دو سری اندازه گیری با استفاده از تجزیه و تحلیل رگرسیون

اگر با آزمون ستیودنت-t وجود اختلاف معنی داری بین دو سری اندازه گیری (برای مثال بین آزمایشگاه روتین و آزمایشگاه کنترلی) نمایان شد، آنگاه آنالیزهای آزمایشگاه روتین یا باید تکرار شود و یا باید تصحیح شود. با توجه به اینکه اعتبار نتایج حاصل از آزمایشگاه کنترلی براساس مطالعات قبلی اثبات شده است، لذا می توان از آن به عنوان مرجع تصحیح داده های آزمایشگاه روتین استفاده کرد.