پردازش آماری داده ها در گزارش آنالیز شیمیایی:

همیشه دو واژه خطا(error) و عدم قطعیت(uncertainty) با نتایج به دست آمده از هرنوع آنالیز و هر نوع اندازه گیری همراه و مطرح می باشد. به بیان دیگر تقریبا محال است اندازه گیری داشته باشیم که این دو واژه در آن مصداق نداشته باشد. به این ترتیب می توان گفت تمامی آنالیزها و اندازه گیری ها یک تخمینی از نتیجه واقعی(true value) در اختیار قرار می دهند.

یکی از راهکارها برای کاهش خطا و افزایش قطعیت(certainty) به نتایج آنالیز انجام تکرار(replicate) برای هر آنالیز می باشد؛ معمولا سه تکرار برای بیشتر انواع آنالیزهای شیمیایی و حتی دستگاهی، تعداد تکرار تایید شده ای می‌باشد.

پارامترها و مفاهیم آماری رایج:

برای یک مجموعه تکرار دو پارامتر آماری میانگین و میانه کاربرد دارد:

- **Average (arithmetic mean, mean): $\bar{X}$: میانگین**

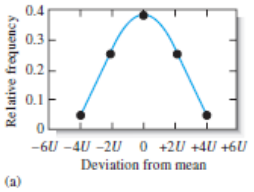
میانگین از طریق تقسیم مجموع نتایج تکرارهای یک آنالیز($X_1,\ldots, X_N$) بر تعداد تکرارها(number of replicates) مطابق فرمول زیر به دست می آید. میانگین تخمینی برای عدد و ارزش میانی(central value) و محتملترین ارزش عددی برای نتیجه آنالیز مورد نظر است.

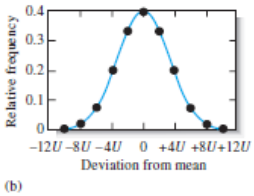
$$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N} \qquad (5\text{-}1)$$

- **Median: میانه**

میانه ارزش و عددی از میان اعداد نتایج تکراهای یک آنالیز است؛ تعداد نتایج کوچکتر و تعداد نتایج بزرگتر از میانه یکسان است.

زمانی می توان به آنالیز و تکرارهای آن اعتماد کرد که نتایج به دست آمده از تکرار یک آنالیز، توزیع(distribution) یکنواخت به صورت منحنی زنگوله ای یا Guasssian مطابق شکل زیر را داشته باشند؛ یعنی تعداد نتایج کوچکتر از میانه و تعداد نتایج بزرگتر از میانه متعادل باشد.





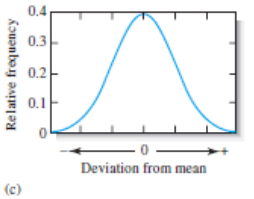


Figure 6-2 Frequency distribution for measurements containing (a) four random uncertainties, (b) ten random uncertainties, and (c) a very large number of random uncertainties.

همچنین مفاهیم آماری زیر در بررسی نتایج یک آنالیز کاربرد دارد:

❖ Precision:دقت

<u>مفهوم واژه precision یا دقت در یک آنالیز</u> این است که نتایج تکرار یک نوع آنالیز به هم نزدیک باشد؛ و در تکرار های بعدی همان آنالیز نیز نتیجه مشابه به دست آید؛ به بیان دیگر قابلیت تکرارپذیری برای نتیجه آنالیز وجود داشته باشد.

سه مفهوم آماری زیر برای توصیف دقت یک آنالیز و تکرارهای آن مطرح و قابل مقایسه می باشد:

Standard Deviation (SD or D): انحراف معیار یا انحراف استاندارد

Variance (واریانس):مربع انحراف استاندارد

Coefficient of variation (CV): ثابت واریانس

❖ Standard Deviation (SD یا d): انحراف استاندارد

انحراف استاندارد(SD) بیانی از مقدار انحراف(deviation) از میانگین می باشد؛ بدیهی است بزرگیSD یا d محدوده فاصله نتایج از میانگین(X) و بنابراین میزان دقت آنالیز را مشخص می کند. عدد انحراف(d) معمولا به صورت قدرمطلق محاسبه می شود؛ و البته می تواند با علامت منفی یا مثبت گزارش شود.

$$d_i = |x_i - \bar{x}| \qquad (5\text{-}2)$$

❖ **Accuracy(صحت):** closeness of measurement to true or accepted value:

<u>مفهوم واژه accuracy یا صحت یک آنالیز</u> نزدیک بودن ارزش نتیجه به دست آمده از آنالیز به ارزش واقعی(true value) و ارزش تعیین شده یا پذیرفته شده(accepted value) برای نمونه مورد آنالیز و آنالیت می باشد.

مقایسه انواع نتایج دقت و صحت کم و زیاد با هدف گذاری بر روی مرکز هدف در شکل زیر مشخص شده است.

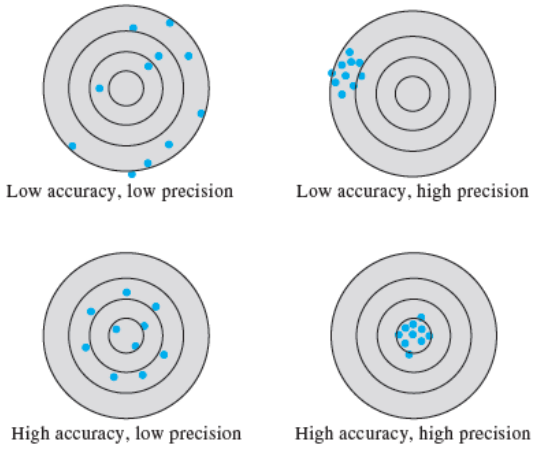


Figure 5-2 Illustration of accuracy and precision using the pattern of darts on a dartboard. Note that we can have very precise results (upper right) with a mean that is not accurate and an accurate mean (lower left) with data points that are imprecise.

در بیان محاوره ای اصطلاح صحت و دقت به طور مترادف به نظر می آید؛ ولي در بيان های علمی این دو از هم متفاوت هستند.

يك نتيجه صحيح و accure آن است كه به مقدار واقعي(true value) خيلي نزديك باشد. به بیانی دیگر اگر نتیجه آنالیز صحیح نباشد، یعنی آنالیز همراه با خطا همراه است. هرچه خطا کمتر باشد، صحت آنالیز بیشتراست.

یافتن مقدار صحیح بودن(صحت) برای نتایج یک آنالیز معمولا سخت است، چون ارزش واقعی معمولا به طور ناشناخته(unknown) می باشد. صحت یک آنالیز با عبارات خطای مطلق(absolute error) و خطای نسبی(relative error) بیان و توصیف می شود.

❖ **Absolute Error(E): خطای مطلق**

خطای مطلق حاصل تفاصل نتیجه یک آنالیز x_i نسبت به ارزش واقعی یا ارزش پذیرفته شده x_t برای آن آنالیت است.

$$E = x_i - x_t \qquad (5\text{-}3)$$

❖ **Relative error(E_r): خطای نسبی**

همانطور که در فرمول پایین مشخص می باشد خطای مطلق نسبت به ارزش واقعی خطای نسبی را نشان می دهد. خطای نسبی کاربردی‌تر از خطای مطلق در گزارش یک آنالیز می باشد.

$$E_r = \frac{x_i - x_t}{x_t} \times 100\% \qquad (5\text{-}4)$$

سوال- در اجرای یک نوع آنالیز خطای نسبی در اشل (scale) کوچک بیشتر می شود یا کمتر؟ و یا قابل پیش بینی نمی-باشد؟

خطا(error):

خطا در شيمي تجزيه عبارت از اختلاف بين نتيجه يک آناليز يا ارزش تعيين شده(accepted value) و ارزش حقيقي(true value) مي باشد.

انواع خطا:

خطا بر سه نوع زير تقسيم مي شود:

1. **Determinate error: systematic errors: affects accuracy**
2. **In-determinate or random error: affects precision**
3. **Gross errors**

1. Determinate or systematic Errors **خطاهاي معين يا سيستماتيک**

خطاهائي که نسبت به روش آناليز انتخابی برای آنالیت مورد نظر در اثر عواملي مشخص، معلوم و قابل پيش بيني و به بيانی به صورت سيستماتيک به وجود مي آيند، خطاهاي معين(Determinate) ناميده مي شوند. معمولا چنين خطاهائي نسبت به ارزش حقيقي، يك جهتي unidirectional هستند؛ و برخلاف خطاهاي نامعين(In-determinate) است که جواب ها بيش از مقدار واقعي و يا کمتر از آن مي‌باشد، خطاهای معين با احتمال مساوي تغيير مي کنند. اين نوع خطاها بر روی صحت(accuracy) نتيجه تاثير دارند.

خطاهاي معين غالبا تکرار پذير و قابل پيش بيني مي باشند: برای مثال اندازه گيري با وزنه هاي غيردقيق (وزنه خورده شده يا زنگ‌زده)، تقسيم بندي نادرست بورت ، ناخالصي در معرف ها، حلاليت قابل ملاحظه رسوب ، واکنش جانبي در تيتراسيون وحرارت دادن نمونه ... خطاهای معين می باشند.

خطاهاي معين شامل دو دسته ديگر از خطاها نيزمی شوند:

خطاي ثابت (Constant Errors) و خطاي متناسب (Proportional Errors):

خطاي ثابت (Constant Errors):

وقتی در يك نوع آناليز همواره يک عدد يا ارزش مطلق به صورت ثابت به عنوان خطا باقی بماند، اين ارزش يک خطای ثابت معرفی می شود.

برای مثال هنگامي که در خوانش حجم پایانی در تیتراسیون اسید و باز که وابسته به تغییر رنگ معرف pH می باشد، همواره ۰/۱ ميلي ليتر بيشتر از حجم تيترانت نقطه پایانی را تعیین کند، این عدد خطای مطلق است. بدیهی است بسته به حجم کل تیترانت لازم برای نقطه پایانی، تاثیر خطای مطلق متفاوت است؛ مثلا اگر حجم کل تیترانت لازم ۱۰ میلی لیتر باشد، مقدار ۰/۱ میلی لیتر ۱۰ درصد بوده و اگر حجم کل تیترانت ۵۰ میلی لیتر باشد، این حجم ۰/۲ درصد می باشد. به این ترتیب در مثال اول تاثر خطای مطلق بیشتر از مثال دوم می باشد.

<u>خطاي متناسب (Proportional Errors):</u>

ارزش مطلق خطای متناسب وابسته و متناسب با اجزای موجود در نمونه، ناخالصي هاي موجود و یا اجزای مداخله گر می باشد. بدیهی است این نوع خطا با حجم نمونه تغییر مي كند ولي خطاي نسبي ثابت مي ماند. برای مثال درتجزیه به روش يدومتري كه جسم اكسيد كننده كلرات است ، چنانچه ناخالصي شامل برمات ويدات هم داشته باشيم، خطاي متناسب رخ ميدهد؛ نمونه هاي با اشل(scale) بزرگتر، ناخالصي بيشتر و خطای متناسب بزرگتری خواهند داشت.

عوامل ایجاد کننده خطاي معين

۱- خطای روش یا متدیک(methodic error): خطای ناشي از روش انتخاب شده برای آنالیز مورد نظر است که متاثر از خصوصیات مواد شیمیائي موجود در روش آنالیز و آنالیت مورد نظر می باشد.

۲- خطای دستگاهي(instrumental error) : ناشي از ابزار و وسائل مورد استفاده در روش آنالیز و اندازه گيري می باشد؛ این خطای دستگاهی می تواند به ماهیت دستگاه مورد استفاده یا در نحوه به کار گیری دستگاه مورد استفاده در آنالیز یا اندازه گیری و یا به تنظیم بودن دستگاه یعنی کالیبراسیون دستگاه مربوط باشد.

۳- خطای فردي(operative error): به مهارت و تجربه شخص آزمایش کننده بستگی دارد.

2. In-determinate or random Errors **خطاهاي نامعين يا تصادفی**

این نوع خطاها که به نامعین یا تصادفی معروف هستند، قابل پیش بینی نبوده و منجر به تنوع در اعداد نتیجه آنالیز در جهت کمتر یا بیشتر می تواند باشد.

3. Gross errors

خطاهای نوع سوم خطاهایی هستند که گاهی می تواند اتفاق بیفتد و گاه اتفاق نیفتد. و تغییر بزرگی در نتیجه در هر یک از دو جهت کاهشی یا افزایشی ایجاد می کند. معمولا این نوع خطا وابسته به بی توجهی فرد آزمایشگر در اجرای آنالیز می باشد؛ مثلا قسمتی از رسوب یا محلول از آنالیز جا می ماند.

روش هاي پذيرش يا رد نتيجه مشكوك(suspected date) در آنالیز

تصمیم گیری نسبت به ظهور یک داده پرت یا outlier در تکرارهای زیاد در مقایسه با داده پرت از تکرار کم از یک آنالیز قابلیت اعتماد بیشتری دارد.

سه روش برای پذیرش یا رد(rejection) داده پرت وجود دارد:

Outlier data: decision rules to confirm or reject suspected data:

- ✓ 2.5 D
- ✓ 4 D
- ✓ Q rejection test

روش 2.5 D یا 4D:

اين روش براي حذف يك نتيجه مشكوك بين ۴ تا ۸ نتيجه بدست آمده كاربرد داشته و بترتيب ذيل محاسبه مي شود:

- ميانگين نتايج قابل پذيرش محاسبه می شود.
- انحراف (d يا D) هر يک از نتايج از ميانگين (تفاضل هر يک از نتايج از ميانگين) محاسبه مي شود.
- ميانگين انحراف ها از ميانگين عددی نتايج نيز محاسبه می شود.
- سپس تفاضل يا انحراف نتيجه مشكوك از ميانگين نتايج قابل پذيرش محاسبه مي شود.

در صورتيكه انحراف نتيجه مشكوك ازميانيگن نتايج قابل پذيرش بيش از ۲/۵ يا ۴ برابر ميانيگن انحراف نتايج قابل پذيرش باشد، نتيجه مشكوك را بايد حذف كنيم؛ درغير اينصورت بايد در نظر گرفته شود.

مثال- چهار محاسبه نرماليته براي يك محلول به ترتيب ذيل بدست آمده است.

0.1014، 0.1102 ، 0.1001، 0.0985

با استفاده از روش D مشخص نمائید که داده ای برای حذف وجود دارد یا خیر؟

روش Q یا Q–test:

این روش براي حذف يك نتيجه مشكوك بترتيب ذيل محاسبه مي شود:

- دامنه نتايج با در نظر گرفتن نتيجه مشكوك محاسبه مي شود (تفاضل بين بيشترين و كمترين نتايج)
- اختلاف بين نتيجه مشكوك و نزديكترين نتيجه محاسبه مي شود.
- براي بدست آوردن ضريب حذف Q(rejection quotient) عدد تفاوت نتيجه مشكوک با نزديكترين نتيجه بر دامنه نتايج تقسيم مي شود.=Q
- با استفاده از جدول حاوي اعداد Q با احتمالات ۹۰ يا ۹۵درصد بر حسب تعداد نمونه آناليز قضاوت صورت مي گيرد: در صورتيكهQ به دست آمده از تقسيم بيش از ارزش Q متناسب از جداول اطلاعات آمار باشد، ميتوان نتيجه را با اطمينان ۹۰يا ۹۵ درصد حذف كرد.