❖ **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 نتیجه تمامی آنالیزها و اندازه گیری ها **تخمینی از نتیجه واقعی(true value)** می باشد.
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 از یک آنالیز، دو پارامتر آماری **میانگین** و **میانه** کاربرد دارد:

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

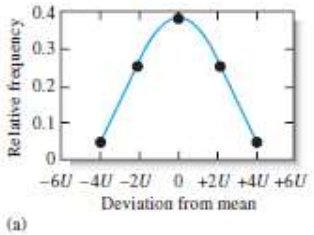
میانگین از تقسیم مجموع نتایج تکرارهای یک آنالیز($X_1,\ldots, X_N$) بر تعداد تکرارها(number of replicates) مطابق فرمول زیر به دست می آید.

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N} \qquad (3\text{-}1)$$

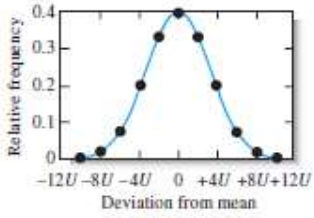
میانگین، یک تخمین برای عدد و ارزش میانی (central value) در بین اعداد نتایج حاصل از آنالیز است؛ میانگین محتملترین ارزش عددی برای نتیجه آنالیز مورد نظر است.

- **Median: میانه**

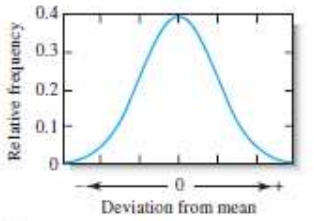
میانه ارزش و عددی از میان اعداد نتایج حاصل از تکرارهای یک آنالیز است؛
تعداد نتایج **کوچکتر** از میانه و تعداد نتایج **بزرگتر** از میانه **یکسان** است.
زمانی می توان به آنالیز و تکرارهای آن اعتماد کرد که نتایج به دست آمده از تکرار یک آنالیز، توزیع(distribution) **یکنواخت** به صورت **منحنی زنگوله ای یا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



(a)



(b)



(c)

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.

- **<u>یک مثال برای تفاوت میانگین (mean) و میانه (median)</u>**:

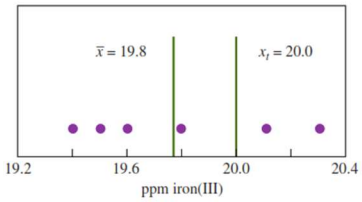


FIGURE 3-1
Results from six replicate determinations of iron in aqueous samples of a standard solution containing 20.0 ppm iron(III). The mean value of 19.78 has been rounded to 19.8 ppm (see Example 3-1).

EXAMPLE 3-1

Calculate the mean and median for the data shown in Figure 3-1.

Solution

$$\text{mean} = \bar{x} = \frac{19.4 + 19.5 + 19.6 + 19.8 + 20.1 + 20.3}{6} = 19.78 \approx 19.8 \text{ ppm Fe}$$

Because the set contains an even number of measurements, the median is the average of the central pair:

$$\text{median} = \frac{19.6 + 19.8}{2} = 19.7 \text{ ppm Fe}$$

در ادامه به مفاهیم آماری با کاربرد رایج در پردازش و مطالعه نتایج یک آنالیز پرداخته می شود:

- **Precision:دقت**

مفهوم کلیدواژه precision (دقت) در یک آنالیز این است که نتایج تکرار یک نوع آنالیز به هم نزدیک باشد؛ و در تکرار های بعدی از آن آنالیز نیز نتیجه مشابه به دست آید؛ به بیان دیگر **قابلیت تکرارپذیری (reproducibility یا repeatability)** برای آنالیز وجود داشته باشد.

سوال- تفاوت بین **reproducibility** و **repeatability** چیست؟

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

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

Variance (واریانس):مربع انحراف استاندارد

Coefficient of variation (CV): ثابت واریانس

❖ **Standard Deviation (SD or D or d): انحراف استاندارد**

انحراف استاندارد(SD) بیانی از مقدار انحراف(deviation) که نتیجه عددی هر یک از آنالیزهای تکرار شده از عدد میانگین حاصل از نتایج آنالیز داشته باشد؛ بدیهی است بزرگیSD یا D یا d محدوده فاصله عددی هر یک از نتایج از میانگین(X) را نشان می دهد.

بنابراین SD میزان دقت آنالیز را مشخص می کند. عدد انحراف معمولا به صورت **قدرمطلق** محاسبه می شود؛ و البته می تواند با **علامت منفی یا مثبت** گزارش شود.

$$d_i = |x_i - \bar{x}| \qquad (3\text{-}2)$$

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

مفهوم کلیدواژه accuracy یا صحت یک آنالیز نزدیک بودن ارزش و بزرگی عددی نتیجه به دست آمده از آنالیز به **ارزش واقعی(true value)** و **ارزش تعیین شده یا پذیرفته شده(accepted value)** برای نمونه مورد آنالیز (آنالیت) می باشد.

در شکل زیر (با طرح دارت و بورد دارت) **دقت** و **صحت** از نظر انواع حالات ممکن (کم low و زیاد high) در بین نتایج حاصل از تکرارهای یک آنالیز با هدف گذاری بر روی مرکز هدف مقایسه شده است.

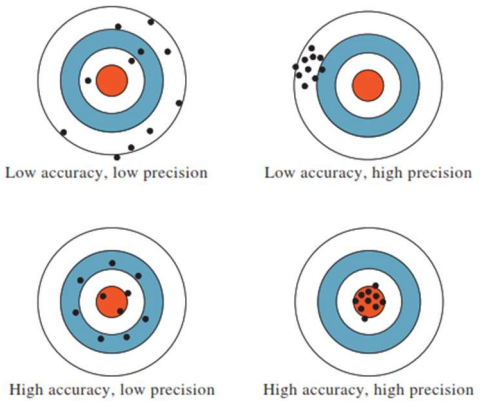


FIGURE 3-2
Illustration of accuracy and precision using the pattern of darts on a dart-board. Note that we can have very precise results (*upper right*) with a mean that is not accurate and an accurate mean (*lower left*) with data points that are imprecise.

در بیان محاوره ای اصطلاح صحت و دقت شاید مترادف به نظر می آید؛ ولي در بيان 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، به معنی آن است که آنالیز با **خطا (error)** همراه بوده است.

هرچه خطا کمتر باشد، صحت آنالیز بیشتر است.

یافتن مقدار صحیح و صحیح بودن(صحت) برای نتایج یک آنالیز معمولا سخت است، چون ارزش واقعی برای یک آنالیت در یک آنالیزمعمولا 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.

- **خطای مطلق :Absolute Error(E)**

خطای مطلق حاصل **تفاضل نتیجه یک آنالیز x_i از ارزش واقعی یا ارزش پذیرفته شده x_t** برای آن آنالیت است.

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

- **خطای نسبی :Relative error(E_r)**

همانطور که در فرمول پایین مشخص می باشد خطای نسبی بر اساس **نسبت خطای مطلق به ارزش واقعی** برای آن آنالیت به دست می آید.

$$E_r = \frac{x_i - x_t}{x_t} \times 100\% \qquad (3\text{-}4)$$

محاسبه خطای نسبی برای مثال قبلی از کتاب اسکوگ وست:

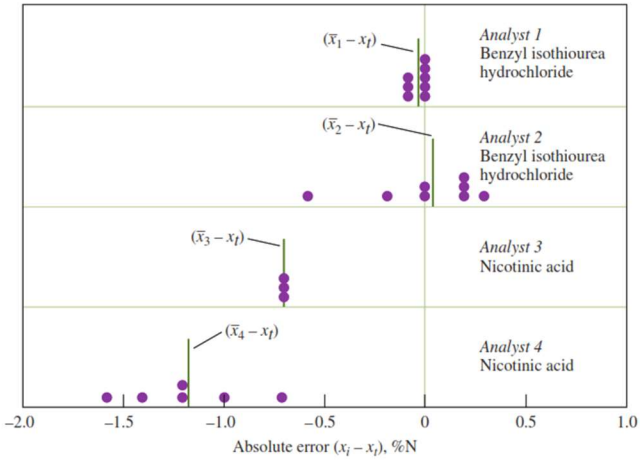
Relative error is also expressed in parts per thousand (ppt). For example, the relative error for the mean of the data in Figure 3-1 is

$$E_r = \frac{19.8 - 20.0}{20.0} \times 100\% = -1\%, \text{ or } -10 \text{ ppt}$$

در گزارش یک آنالیز **خطای نسبی کاربردی‌تر از خطای مطلق** می باشد.

سوال- در اجرای یک آنالیز در اشل یا ابعاد (scale) کوچک خطای نسبی ... می شود. بیشتر یا کمترو یا قابل پیش بینی نمی‌باشد؟

سوال- نتایج عددی حاصل از آنالیز مقدار نیتروژن در دو ترکیب بنزیل ایزوتیواوره (benzyl iso-thiourea) و نیکوتین اسید (نیاسین یا ویتامین B3) با ساختارهای رسم شده در زیربه عنوان آنالیت با ۴ آنالیزگر مختلف (analyst) در نمودار زیر آمده است. نتایج را بر اساس تفاوت در دقت و صحت مقایسه و تفسیر نمایید.



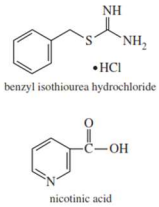


FIGURE 3-3
Absolute error in the micro-Kjeldahl determination of nitrogen. Each dot represents the error associated with a single determination. Each vertical line labeled $(\bar{x}_i - x_t)$ is the absolute average deviation of the set from the true value. (Data from C. O. Willits and C. L. Ogg, *J. Assoc. Offic. Anal. Chem.*, **1949**, *32*, 561.)

❖ **خطا(error):**

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

❖ **انواع خطا:**

خطا بر سه نوع زير تقسيم بندی 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 با توجه به روش آناليز (analysis method) انتخاب شده برای آناليت مورد نظر در اثر عواملي **مشخص، معلوم و قابل پيش بينی و به بيانی به صورت سيستماتيک و اجتناب ناپذير** به وجود مي آيند، خطاهاي معين(determinate) ناميده مي شوند.

معمولا خطاهاي معين نسبت به ارزش حقيقي (مقدار واقعی) يك جهتي unidirectional بوده و با احتمال مساوي تغيير مي کنند. اين نوع خطاها بر روی صحت نتيجه(accuracy of result) تاثير دارند. خطاهاي معين به صورت غالب تکرار پذير و قابل پيش بيني مي باشند: برای مثال اندازه گيري با وزنه هاي غيردقيق (وزنه خورده شده يا زنگـزده)، درجه بندی نادرست بورت، ناخالصي يا ناپايداری شيميايی در معرف ها، حلاليت قابل ملاحظه رسوب در مسير يک آناليز، واکنش جانبي در تيتراسيون و تغييرات حاصل از حرارت دادن نمونه ... خطاهای معين می باشند.

❖ **منابع ايجاد کننده و انواع خطاي معين**

خطای روش (method error)

خطای ابزار يا دستگاه(instrumental error):

خطای فردی يا آزمايشگر(personal error)

- **خطای روش (method error):**

خطای ناشي از روش انتخاب شده برای آنالیز مورد نظر می تواند باعث بروز خطای معین در یک آنالیز شود. خطای روش متاثر از مراحل روش آنالیز، خصوصیات مواد شیمیائي مورد استفاده در روش آنالیز و خصوصیات شیمیایی آنالیت مورد نظر در روش آنایز به کار گرفته شده می باشد.

این نوع خطا با اجرای آنالیز بر روی نمونه های رفرانس (مرجع) استاندارد ((Standard Reference Material (SRM) (در قالب استاندارد خارجی یا استاندارد داخلی)، استفاده از نمونه بلانک و نیز مقایسه نتایج با یک روش آنالیز متفاوت کاهش و یا حذف می شود.

سوال- استاندارد خارجی چیست؟

سوال- استاندارد داخلی چیست؟

سوال- نمونه بلانک چه ویژگی هایی دارد؟

- **خطای ابزار یا دستگاه(instrumental error):**

خطای ناشي از ابزار و وسائل مورد استفاده در روش آنالیز و اندازه گیري می تواند باعث بروز خطای شود. خطای دستگاهی می تواند به ماهیت دستگاه مورد استفاده، یا نحوه به کار گیری دستگاه مورد استفاده در مراحل آنالیز (دستگاه مورد استفاده در اندازه گیری و ...) و یا به تنظیم بودن دستگاه یعنی کالیبراسیون دستگاه مربوط باشد.

این نوع خطا با کالیبراسیون دستگاه در زمان های منظم قبل از هر آنالیز کاهش و یا حذف می شود.

سوال- کالیبراسیون دستگاه چیست؟

- **خطای فردی یا آزمایشگر (personal error):**

این نوع خطا بر اساس سطح مهارت، دقت و البته تجربه شخص آزمایش کننده (analyst) در روش آنالیز اجرا شده دیده می شود.

این نوع خطا با رعایت نظم و مسیر اجرای استاندارد Standard Operating Process (SOP) کاهش و یا حذف می شود.

❖ خطاهاي معين در دو دسته از خطاها به شرح زیر می توانند معرفی شوند:

خطاي معين از دسته خطای ثابت (Constant Errors)

خطاي معين از دسته خطای متناسب (Proportional Errors)

- **خطاي معين از دسته خطای ثابت (Constant Errors)**:

وقتی در يك نوع آناليز با ابعاد (scale) مختلف، خطا همواره يک عدد يا ارزش مطلق به صورت ثابت باقی بماند، اين ارزش يک خطای ثابت معرفی می شود. بديهی است خطای ثابت در آناليز با ابعاد کوچکتر مهمتر و قابل توجه تر از آناليز با ابعاد بزرگتر می باشد. در حضور خطای ثابت در يک آناليز خطای مطلق با تغيير در ابعاد (scale) آناليز انجام شده ثابت می ماند اما خطای نسبی با تغيير در ابعاد (scale) آناليز انجام شده تغيير می کند.

- **خطاي معين از دسته خطای متناسب (Proportional Errors)**

ارزش مطلق خطای متناسب (proportional) وابسته و متناسب با اجزای موجود در نمونه، ناخالصي هاي موجود و يا اجزای مداخله گر در نمونه مورد آناليز می باشد. بديهی است اين نوع خطا با حجم نمونه تغيير مي کند ولي خطاي نسبي ثابت مي ماند. برای مثال درتجزيه به روش يدومتري که جسم اکسيد کننده کلرات است ، چنانچه ناخالصي شامل برمات ويدات هم داشته باشيم، خطاي متناسب رخ ميدهد؛ دراين مثال نمونه هاي با ابعاد (scale) بزرگتر، به دليل ناخالصي بيشتر، خطای متناسب بزرگتری خواهند داشت.

2. In-determinate or random Error

خطاهاي نامعين يا تصادفی

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

خطاهاي نامعين(In-determinate) در دو جهت (bi-directional) شامل بیشتر یا کمتر از ارزش حقیقی می توانند مشاهده شوند.

3. Gross errors

خطاهای نوع سوم خطاهایی هستند که گاه پیش می آید و گاه پیش نمی آید. این نوع خطا تغییر بزرگی در نتیجه یک آنالیز در هر یک از دو جهت کمتر یا بیشتر می تواند ایجاد کند. معمولا این نوع خطا وابسته به بی توجهی فرد آزمایشگر در اجرای آنالیز می باشد؛ مثلا قسمتی از رسوب یا محلول از آنالیز در مسیر روش آنالیز از دست رفته، به کار گرفته نشود.

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

❖ **قوانین تصمیم گیری به پذیرش یا رد نتیجه مشکوك(suspected date)یا داده پرت (outlier data) در آنالیز:**

در میان نتایج عددی حاصل از تکرارهای یک آنالیز بر روی یک نمونه در صورتی که به کاهش یا حذف احتمال خطاهای معین (از هر سه نوع خطای معین شناخته شده) توجه شده باشد ممکن است به دلیل خطاهای gross ناشناخته ای یک یا چند داده به دست آمده از بقیه داده های حاصل فاصله عددی قابل توجه داشته باشند. این داده ها داده پرت یا outlier و یا داده مشکوک (suspected) نامیده می شود. در اینجا می توان از قوانین آماری برای تصمیم گیری به حفظ (retention) یا حذف (rejection) داده پرت یا outlier استفاده برد. این قوانین در نرم افزارهای آماری به صورت برنامه و تابع مشخص نیز تعریف شده می باشد.

بدیهی است تصمیم گیری برای حفظ یا حذف یک داده پرت در یک آنالیز با تعداد تکرارهای (replicate) بیشتر در مقایسه با همان آنالیز با تعداد تکرار کمتر قابلیت اعتماد بیشتری دارد.

سه روش برای حفظ یا حذف (rejection) داده پرت یا داده مشکوک وجود دارد:

- **2.5 D or 4 D**
- **Q rejection test**

- **روش 2.5 D یا 4D:**

این روش براي حذف يك داده مشكوك بين ۴ تا ۸ نتيجه به دست آمده از آنالیز کاربرد داشته و به ترتیب زیر محاسبه مي شود:

۱- ابتدا باید داده های قابل حفظ یا پذیرش از داده های مشکوک جدا شود.

۲- میانگین **داده های قابل حفظ یا پذیرش** محاسبه شود: $\overline{X}$

۳- در ادامه انحراف استاندارد برای **داده های قابل پذیرش** نیز محاسبه شود: $D = X_i - \overline{X}$

۴- سپس میانگین انحراف استانداردهای محاسبه شده در ردیف ۳ محاسبه شود: $\overline{D}$

۵- همچنین انحراف استاندارد برای هر یک از داده های مشکوک از میانگین داده های قابل پذیرش (محاسبه شده در ردیف۱) محاسبه شود: D'

۶- نسبت انحراف هر یک از داده های مشکوک از میانگین نتایج قابل پذیرش (به دست آمده در ردیف ۵: D') به میانگین انحراف استاندارد داده های قابل پذیرش (به دست آمده در ردیف۴: $\overline{D}$) محاسبه شود: $\frac{D'}{\overline{D}}$

۷- در صورتیکه انحراف هر یک از داده های مشکوك ازمیانیگن نتایج قابل پذیرش بیش از ۲/۵ برابر یا ۴ برابر میانیگن انحراف استاندارد داده های قابل پذیرش باشد (حاصل محاسبه شده در ردیف ۶)، داده مشکوك باید حذف شود؛ درغیر اینصورت باید حفظ شود.

❖ **روش Q یا Q–test:**

این روش براي حذف يك داده مشكوك به ترتيب زير محاسبه مي شود:

۱- دامنه داده های حاصل با در نظر گرفتن داده های مشکوك محاسبه شود: تفاضل بین بیشترین و کمترین نتایج

۲- همچنین اختلاف بین داده مشكوك و نزديكترين داده به داده مشترک مورد تصمیم گیری محاسبه شود.

۳- سپس عدد تفاوت داده مشکوک با نزدیکترین داده (به دست آمده در ردیف ۲) بر دامنه داده ها (به دست آمده در ردیف ۱) تقسيم مي شود؛ اين عدد ضريب حذف Q(rejection quotient) نامیده می شود.

۴- در انتها با استفاده از جدول آمار اعداد Q با احتمالات ۹۰ درصد یا ۹۵درصد (بسته به تصمیم تحقیق) که بر حسب تعداد تکرار آنالیز موجود است، تصمیم گیری صورت مي گيرد:

در صورتیکه Q به دست آمده در ردیف ۳ بیش از ارزش Q متناسب در جداول آمار اعداد Q باشد، بر اساس این قانون داده با اطمینان ۹۰ درصد یا ۹۵ درصد باید حذف شود.

سوال- در یک تیتراسیون تعیین نرمالیته اسید با چهار تکرار چهار داده زیر با چهار رقم اعشار به دست آمده است.

0.1014

0.1102

0.1001

0.0985

با استفاده از هر یک از دو قانون تصمیم گیری گفته شده مشخص نمائید که آیا داده پرت یا قابل حذف در میان این داده ها وجود دارد یا خیر؟