Chemometrics

Application of mathematical, statistical, graphical or symbolic methods to maximize chemical information.

-However, this definition can be expanded to include:

biology (biometrics),

environmental science (environmetrics),

economics (econometrics).

-Two lines of development:

- experimental design: planning and performing experiments in a way that the resulting data contains the maximum information about stated questions.
- multivariate data analysis: utilizing all available data in the best possible way.

1

Chemometrics

1. Errors in Quantitative Analysis & Descriptive Statistics
2. Significance Tests & Computer Exercise
3. Regression & Correlation Methods, Calibration Methods & Computer Exercise
4. Non-Parametric & Robust Methods
5. Experimental Design Technique
6. Multivariate Methods of Analysis & Computer Exercise
7. Quality Assurance & Good Laboratory Practice

2

Errors in Quantitative Analysis

- Student A

  Results have two characteristics:

  1. All very close to each other (10.08-10.12 mL).
  2. All too high (10.00 mL exactly).

  Two types of errors have occurred:

  1. Random – cause replicate results to differ from one another → results fall on both sides of the mean (10.10 mL in student's A case).
     → effect the precision or reproducibility of the experiment.
     → Student A – small random errors (precise).
  2. Systematic – cause all the results to be in error in the same sense (High).
     → Total systematic error is termed bias.

  Hence: Student A has precise, but biased results.

3

Errors in Quantitative Analysis

- No analysis is free of error!
- Types of errors
  1. Gross – readily described, errors that are obvious.
     Instrument breakdown, dropping a sample, contamination (gross).
  2. Random
  3. Systematic

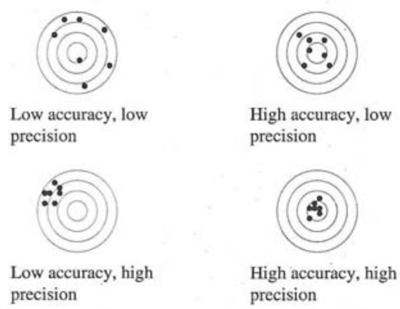
Student	Results (mL)					Comments
A	10.08	10.11	10.09	10.10	10.12	Precise, biased
B	9.88	10.14	10.02	9.80	10.21	Imprecise, unbiased
C	10.19	9.79	9.69	10.05	9.78	Imprecise, biased
D	10.04	9.18	10.02	9.97	10.04	Precise, unbiased

Titration – each student performs an analysis in which exactly 10.00 mL of exactly 0.1 M NaOH is titrated with exactly 0.1M HCl.

4

Errors in Quantitative Analysis

- Accuracy – How far a result is from the true value.
- Precision – How close multiple determinations are to each other.



5

Handling Systematic Errors

1. Identifying Problem Areas – before the experiment starts
2. Careful Experimental Design
3. Instrument Performance – checks, maintenance
4. Use of Standard Reference Materials (SRMs)
5. Intercomparison Methods
6. Proficiency Testing Schemes

6

Exercises

1. See Problem #1, page 16

   Comment on the bias, precision and accuracy of each of these set of results.
3. The number of binding sites per molecule in a sample of a monoclonal antibody is determined four times with results of 1.95, 1.95, 1.92, and 1.97. Comment of the bias precision and accuracy of these results.
4. Discuss the degrees of bias and precision desirable or acceptable in the following analysis:

   a. Determination of the lactate concentration of human blood samples.
   b. Determination of uranium in an ore sample.
   c. Determination of a drug in blood plasma after overdose.
   d. Study of the stability of a calorimetric reagent by determination of its absorbance at a single wavelength over a period of several weeks.

Introduction to JPM Chemometric Software

1. SPSS based approach
2. Experimental Design
3. Multivariate Analysis
4. Routine Statistical Analysis including:

   a. Descriptive stats
   b. Calibration
   c. ANOVA analysis

5. Allows us to see trends in data, important factors, and their interactions.

Descriptive Statistics

1. Basic Statistics ⟶ Descriptive

Spectrophotometric measurement (Abs) of a sample solution from 15 replicate measurements.

Measurement	Value	Measurement	Value
1	0.3410	9	0.3430
2	0.3350	10	0.3420
3	0.3470	11	0.3560
4	0.3590	12	0.3500
5	0.3530	13	0.3630
6	0.3460	14	0.3530
7	0.3470	15	0.3480
8	0.3460		

Descriptive Statistics

Descriptive statistics for the spectrophotometric measurements.

Parameter	**Value**
Sample #, n	15
Mean	0.3486
Median	0.347
Std Dev	0.00731
RSD %	2.096
Std error	0.00189
Max value	0.363
Min value	0.335

Statistical Tests – Student's *t*-test, *F*-test, tests for outliers
Distributions – Gaussian, Poisson, binominal

Descriptive Statistics

1. Statistics of Repeated Measurements

$$Mean\ \bar{x} = \frac{\sum x_i}{n}$$

Student A $\quad Standard\ deviation\ \ s = \sqrt{\sum (x-\bar{x})^2/(n-1)}$

	x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
	10.08	-0.02	0.0004
	10.11	0.01	0.0001
	10.09	-0.01	0.0001
	10.10	0.00	0.0000
	10.12	0.02	0.0010
Total	50.50	0	0.0010

11

Descriptive Statistics

$$\bar{x} = \frac{\sum x_i}{n} = \frac{50.50}{5} = 10.1\,mL$$

$$s = \sqrt{\sum (x-\bar{x})^2/(n-1)} = \sqrt{0.001/4} = 0.0158\,mL$$

12

My other job!

13

Descriptive Statistics

2. Distribution of Repeated measurements
   1. Gaussian distribution
      - **Bell-shaped curve for the frequency of the measurements**

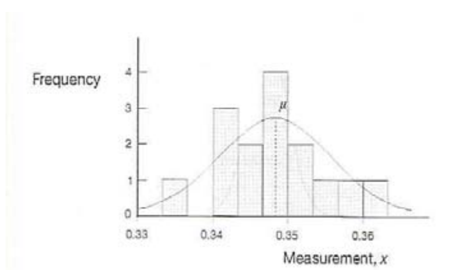


Fig. 1 – Histogram for the measurements of spectrophotometric data. Theoretical distribution with the Gaussian curve in the solid line.

14

Descriptive Statistics

A. Gaussian Distribution

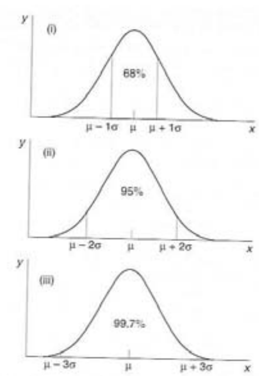


Figure 2.4 Properties of the normal distribution: (i) approximately 68% of values lie within $\pm 1\sigma$ of the mean; (ii) approximately 95% of values lie within $\pm 2\sigma$ of the mean; (iii) approximately 99.7% of values lie within $\pm 3\sigma$ of the mean.

15

Descriptive Statistics

Sample – statistical sense of a group of objects or subjects

e.g. sample of 50 measurements of phosphate ion concentration from an infinite population of all such possible measurements.

For a sample of *h* measurements:

std error of the mean = $\sigma/\sqrt{n}$

– larger *n* = ?

→ sampling distribution of the mean

– normal distribution as *n* increases

16

Descriptive Statistics

Sample to define a range – confidence interval

→ range which reasonably assumes inclusion of the true value

Confidence Limits – extreme values of the range

For large samples, the confidence limits of the mean are given by:

$$\bar{x} \pm zs/\sqrt{n}$$

where the value of z depends on the degree of confidence required:

for 95.1 C.L., $z = 1.96$
for 99.1 C.L., $z = 2.58$
for 99.7 C.L., $z = 2.97$

17

Descriptive Statistics

Confidence Limits of the mean – small samples

$$\bar{x} \pm t_{n-1}\, s/\sqrt{n}$$

(n-1) indicates that t depends on this quantity (degrees of freedom).

# of independent deviations $(x_i - \bar{x})$ which are used in calculating s.

when (n-1) deviations are known, the last can be deduced since $\Sigma(x_i - x) = 0.$ ‾

18