CHE 392
CHEMICAL ENGINEERING LAB I

DATA ANALYSIS

Prof. Dr. N. Alper TAPAN
Assoc. Prof. Dr. Dolunay ESLEK KOYUNCU
Assoc. Prof. Dr. H. Mehmet TAŞDEMİR

OUTLINE

- SIGNIFICANT FIGURES IN EXPERIMENTAL DATA
- STATISTICAL ANALYSIS OF EXPERIMENTAL DATA
- TYPES OF ERRORS
- DEGREE OF FREEDOM, CONFIDENCE INTERVAL
- REFERENCES

SIGNIFICANT FIGURES

- Significant figures are digits showing a numerical value with proper accuracy and precision.

- For instance ,The number of significant figures in 4521 and 6,784 are 4.

- The number of significant figures in 0,006784 and 4521000 are also 4 . The number of zeros before 6 or after 1 are not related with precision .

Some rules about significant figures

- The number of significant figures in 4521000. are 7 because the dot (.) shows that the precision will change by the number following the dot (.)
- Rounding of numbers and change in significant figures;

  52.6502 → 52.7, 3.4577 → 3.46 , 0.34648 → 0.346

  73.135 → 73.14 , 48.724 → 48.72

OPERATIONS IN SIGNIFICANT FIGURES

- Addition and subraction: The precision is based on lowest number of significant figures.

- 32.7 +3.62+10.008=46.328 → 46.3

Significant figures 3 Significant figures 5 Significant figures 3

- Multiplication and division: Like addition or subraction , the result is rounded based on the lowest number of significant figures.

STATISTICAL ANALYSIS OF EXPERIMENTAL DATA

Some important statistical terms;

- Aritmetic Average, $\bar{x}$: division of sum of measurements to number of measurements

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N}$$

x_i: Value of each measurement .

N : # of measurements

STATISTICAL ANALYSIS OF EXPERIMENTAL DATA

- Median: The value in the middle when the data is sorted (ranked)from lowest to highest.

- Mode: The most repating value in the data.
- Range: the difference between the highest and the lowest value in the data.

- Accuracy: the difference between the calculated average experimental value $(\bar{x})$ and the exact value "μ" $(\bar{x} - \mu)$.

Accuracy can be shown as absolute or relative.

STATISTICAL ANALYSIS OF EXPERIMENTAL DATA

- For instance, the exact value of an experimental datum is μ= % 0,54 and by the repetitive experiments the average value is

$\bar{x}$= %0,49

Absolute accuracy= % 0,49-% 0,54= -% 0,05

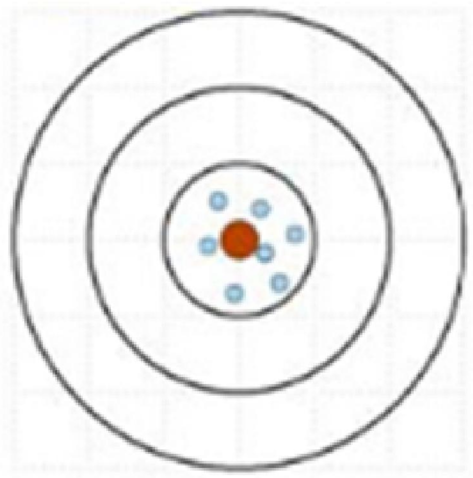
Relative accuracy= (-% 0,05)/(% 0,54)= -% 0,09

STATISTICAL ANALYSIS OF EXPERIMENTAL DATA

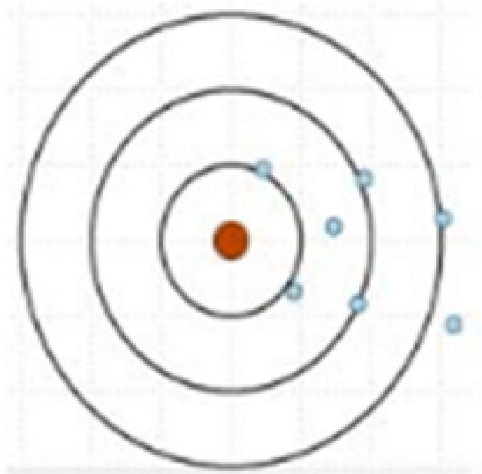
- Certainity: The proximity of the experimental values; Certainity can be determined by;

  - ✓ Standard deviation
  - ✓ Relative standard deviation
  - ✓ Variance
  - ✓ Range
  - ✓ Error
  - ✓ Average error
  - ✓ Relative average error

?
Accuracy = Certainity



High Accuracy
High Certainity



Low Accuracy
Low Certainity

STATISTICAL ANALYSIS OF EXPERIMENTAL DATA

- Standard deviation: Distribution of experimental data arround average value. The smaller the standard deviation higher the certainity. But this does not mean high accuracy. Most frequently used types of standard deviations;
  1. Population standard deviation (σ)
  2. Sample standard deviation (s)
  3. Composite standard deviation ($s_{composite}$)

STATISTICAL ANALYSIS OF EXPERIMENTAL DATA

- Population standard deviation;

$$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{N}}$$

- Sample standard deviation(s)

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{N-1}}$$

Örnek 3.4. Doğal bir su numunesinde klorür tayin ediliyor ve şu d
ğerler bulunuyor.

38,3; 41,1; 35,0; 41,6; 35,4; 36,6 (mg/L)

Buna göre, sudaki klorür konsantrasyonunu, standart sapmasını
ve variyansı (s^2) hesaplayınız. Bunları hesaplamak için, analiz sonuçları
dan bir tablo hazırlanır.

Analiz sonuçları x_i	Ortalamadan sapmalar $(x_i - \bar{x})$	Ortalamadan sapmaların kareleri (fark kareleri) $(x_i - \bar{x})^2$
38,3	0,3	0,09
41,1	3,1	9,61
35,0	-3,0	9,00
41,6	3,6	12,96
35,4	-2,6	6,76
36,6	-1,4	1,96
Toplam : 228,0	0,0	40,38 (fark kareleri toplam

Bu tablodan ortalama değer,

$$\bar{x} = \sum x_i / n$$

$$\bar{x} = 288/6 = 38,0 \text{ mg/L}$$

numune variyansı,

$$s^2 = \sum (x_i - \bar{x})^2 / (n - 1)$$

$$s^2 = 40,38/5 = 8,08 \text{ (mg/L)}^2$$

numune standart sapması (s),

$$s = \text{Variyansın kare kökü} = \sqrt{\sum (x_i - \bar{x})^2 / (n - 1)}$$

$$s = \sqrt{8,08} = 2,84$$

$$s = 2,84 \text{ mg/L}$$

bulunur.

STATISTICAL ANALYSIS OF EXPERIMENTAL DATA

- Composite standard deviation:

$$s_{composite} = \sqrt{\frac{\sum(x_i - \bar{x}_1)^2 + \sum(x_i - \bar{x}_2)^2 + \ldots\ldots + \sum(x_i - \bar{x}_k)^2}{N - k}}$$

- $\bar{x}_1$, $\bar{x}_2$, $\bar{x}_k$: Average values of experiments by different analyzers from different data sets.
- k: # of analyzers

STATISTICAL ANALYSIS OF EXPERIMENTAL DATA

Example:

After fishing , 5 samples of fishes were dried and analyzed by atomic absorbtion spectroscopy (AAS) for mercury determination. Based on this experimental data determine the composite standard deviation.

Analizci	Deney sayısı	Bulunan Hg (ppm)	Ortalama Hg($\overline{x}$)	$\sum(x-\overline{x})$
1	4	1,58; 1,72; 1,86; 1,64	1,70	0,0440
2	5	0,94; 0,88; 0,99; 1,12; 1,06	1,00	0,0361
3	6	3,20; 3,42; 3,16; 3,14; 3,92; 3,08	3,15	0,1334
4	2	2,18; 2,42	2,30	0,0288
5	6	0,58; 0,65; 0,72; 0,48; 0,82; 0,56	0,64	0,0745

Örnekte N = 23, k = 5 dir. Buna göre s, veya s_h,

$$s_h = s = \sqrt{\frac{0,0440 + 0,0361 + 0,1334 + 0,0288 + 0,0745}{23 - 5}}$$

s = 0,13 ppmHg dır.

STATISTICAL ANALYSIS OF EXPERIMENTAL DATA

- Variance: Square of standard deviation. In scientific studies variance is preferred for the expression of certainity.

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{N-1}$$

- Deviation: Difference between each experimental value and average value (d).

$$d = |x_i - \bar{x}|$$

STATISTICAL ANALYSIS OF EXPERIMENTAL DATA

- Average deviation: Division of sum of deviations by the number of experiments.

$$\bar{d} = \frac{(x_1 - \bar{x}) + (x_2 - \bar{x}) + \cdots (x_N - \bar{x})}{N}$$

- Relative average deviation: Division of average deviation by average value ($\bar{x}$).

TYPE OF ERRORS

1. Systematic errors (can be avoided)

- Instrumental
- Methodological
- Personal

2. Random (can not be avoided)

TYPE OF ERRORS

SYSTEMATIC ERRORS

- **Instrumental errors;**

Bad calibration of instruments and unstabilities on power sources.

For instance , during titration the real values may be different from the readings on the burette or graduated pipette, flasks.

RANDOM ERRORS

- Errors that cannot be avoided.
- Independent of the experience of the experimenter. Because of this, there are always some minor errors.
- Random errors may be caused by experimenter or may be due to instrumentation or environmental conditions.
- If we don't know the exact value of error, we determine a range of experimental values by standard deviation or confidence interval.

SOURCES OF TYPE OF ERRORS

Non ideal chemical and physical behaviour of reactants or reactions during analysis cause errors.

Personnal errors are caused to the experimenter . For instance, inadequate washing of precipitate or false recording of data etc.

DEGREE OF FREEDOM, CONFIDENCE LIMITS AND INTERVALS

- A confidence interval (CI) is a range of values that’s likely to include a population value with a certain degree of confidence. It is often expressed a % whereby a population means lies between an upper and lower interval.
- A confidence interval, in statistics, refers to the probability that a population parameter will fall between two set values for a certain proportion of times. Confidence intervals measure the degree of uncertainty or certainty in a sampling method. A confidence interval can take any number of probabilities, with the most common being a 95% or 99% confidence level.
- ***Confidence level*** refers to the percentage of probability, or certainty, that the confidence interval would contain the true population parameter when you draw a random sample many times.

Q test method can be used to discard or accept a suspected experimental value. In this method , different confidence levels can be used , Q test is based on the selected confidence level.

Table . Q values at different confidence levels

Number of experiments	Q%90	Q%96	Q%99
3	0.94	0.98	0.99
4	0.76	0.85	0.93
5	0.64	0.73	0.82
6	0.56	0.64	0.74
7	0.51	0.59	0.68
8	0.47	0.54	0.63
9	0.44	0.51	0.6
10	0.41	0.48	0.57

DEGREE OF FREEDOM, CONFIDENCE LIMITS AND INTERVALS

Example: For the data shown below:
12.53, 12.56 , 12.47 , 12.67 , 12.48 , is 12.67 OK for 90% confidence interval ?

- 12.47, 12.48 , 12.53, 12.56, ← difference = 0.11 → 12.67

  Range= 0.20

- Q = difference / range = 0.55 < $Q_{(\text{value in tablelo})}$ (keep the value)

- If $Q_{(\text{observed})} > Q_{(\text{value in the table})}$ value is thrown based on the confidence level given.

Some significance tests

- Comparison of average ($\bar{x}$) and exact value(μ)
- F-test: Comparison of two certainity
- T-tests
- G-tests: Grubbs tests
- C-tests: Cochran tests

References

- Harris, D.C., “Analytical Chemistry”, W.H. Freeman and Company, 42 , 1982.
- Kreyzig, E., “Advanced Engineering Mathematics” , 8th edition, John Wiley and Sons Inc., 914 , 1999.
- Bennett, C.A., Franklin, N.L., “Statistical Analysis in Chemistry and the Chemical Industry”, John Wiley and Sons Inc., 1967.
- Mickley, H.S., Sherwood, T.K., and Reed, C.E., “Applied Mathematics in Chemical Engineering”, McGraw-Hill, 1957.
- Reilly, P.M., “A Statistical Look at Significant Figures”, Chemical Engineering Education”, summer issue, 152-155,1992.
- Schwatz, L.M., “Propagation of Significant Figures”, Journal of Chemical Engineering Education., 62(8) , 693 , 1985.
- Garland, C.W., Nibler, J.W., “Experiments in Physical Chemistry” ,7th edition, McGraw-Hill, 2003
- Usanmaz, A., “Quantitative Analytical Chemistry”, Middle East Technical University, 1991.
- Gündüz, T., ‘’Kimyacılar için İstatistik”, 2. Baskı, Gazi Kitap evi, 2010
- Skoog, D.A., West, D.M., Holler, F.J., ‘’Analitik Kimya Temelleri’’, Editör: E.Kılıç, F. Köseoğlu, 7. Baskı, Bilim Yayıncılık, 1997.