

Data Analysis II

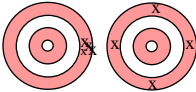
CU- Boulder
CHEM-4181
Instrumental Analysis Laboratory

Prof. Jose-Luis Jimenez
Spring 2007

Lecture will be posted on course web page – based on lab manual, Skoog, web links

2

Summary of Last Lecture

- *Treat data in your lab reports and student choice exp. in a professional way*
- Topics covered in lecture I
 - Significant figures
 - Precision vs. accuracy 
 - Errors
 - E_a and RE
 - Gross errors -> outliers
 - Random errors
 - Treat with statistics (Gaussian distribution)
 - Systematic errors
 - Identify and get rid of them
- Today: treatment of random errors & Excel

3

Population and Sample Mean

- Sample Mean (\bar{x})

- Average of a finite set of data
- In general not the same as μ , because of finite error
- AVERAGE() in Excel

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

- Population Mean (μ)

- Also “limiting mean”
- It is the true value of the quantity being measured

$$\mu = \lim_{N \rightarrow \infty} \frac{\sum_{i=1}^N x_i}{N}$$

4

Standard Deviation and Variance I

- Population Standard Deviation (σ)

- Measure of the precision of a population of data
- STDEVP() in Excel

- Variance (σ^2)

- Std. dev. has same units of x , variance as units of x^2
- Variance from different effects is often additive
 - $\sigma^2 = \sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \dots$ (INDEPENDENT effects)
 - Std. Dev. is not!
- VAR() in Excel

$$\sigma = \sqrt{\lim_{N \rightarrow \infty} \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}}$$

5

Standard Deviation and Variance II

- Sample Standard Deviation (s)
 - s instead of σ
 - \bar{x} instead of μ
 - $(N-1)$ instead of N
 - “Number of degrees of freedom”, $v = N-1$
 - Because \bar{x} is used in the calculation, only $N-1$ values are independent, the last one can be calculated from the mean and the other values
 - STDEV() in Excel

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

6

RSD and CV

- Relative Standard Deviation (RSD)
 - Often more informative than absolute SDs
- $$RSD = \frac{s}{\bar{x}} \cdot 10^z$$
- $z = 2 \Rightarrow$ percent
 - $z = 3 \Rightarrow$ ppth
- Coefficient of Variation (CV)
 - RSD expressed as a percent

$$CV = \frac{s}{\bar{x}} \cdot 10^2$$

7

Standard Error of the Mean

- Standard deviation
 - estimate of the probable error of a single measurement
- Standard error of the mean
 - Estimate of the probable error of the mean of N measurements

$$\sigma_m = \frac{\sigma}{\sqrt{N}}$$

$$s_m = \frac{s}{\sqrt{N}}$$

- More generally
 - The mean of N measurements has a distribution $N(\mu, \sigma_m^2)$
 - This is true in the limit even if error is NOT Gaussian
 - “Central limit theorem” of probability

8

Excel Tutorial – Part 1

Stats Tutorial - Microsoft Excel™ Basics:

Entering Data
– Referencing, pasting and creating data series and navigating cells

Formulas
– Using Excel formulas in cells to enter equations and manipulate data

Plotting & Graphs
– Plotting data using different charts, including an example of [proper graph format](#)

Functions
– Using the built-in functions in Excel to manipulate your data

Adding a Trendline
– Using the trendline function to add a best-fit line to your graph

This section introduces data manipulation using Microsoft Excel™, including importing, copying and pasting data and entering equations. A basic understanding of computer operating systems (Windows/Mac) is assumed, including the ability to launch applications, and find and open files.

Note: different text formats indicate different actions or operations. Things you need to type will be denoted with the Courier font. Menus will be denoted with **bold Arial**, with the menu item in *italic*, e.g. **File** -New means select the New item in the File menu. Options, buttons in dialog boxes and the Enter key will also use this font. Dialog box titles will be in **bold Times**. Functions will be denoted with **ALL-CAPS COURIER**.

To begin, continue to the next page to learn how data is organized in Excel, and how to [enter data](#).

© 2006 Dr. David C. Stone & Ann E.lli, Chemistry, University of Toronto
Last updated: August 29th, 2006

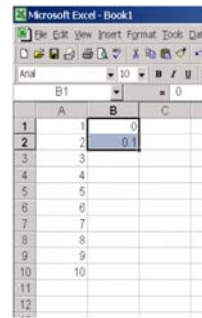
<http://www.chem.utoronto.ca/coursenotes/anslci/StatsTutorial/ExcelBasics.html> 9

Entering Data

- In Excel data are entered in cells
 - Cells can be empty or contain data or formulas
 - Every cell has coordinates
 - A1, B12, etc.
 - Absolute coordinates: \$A1 or A\$1 or \$A\$1
 - Very important distinctions!

CQ: Do you know how to use relative and absolute references in Excel?

A. yes
B. a little
C. no



- Demo: pasting a series of data
 - Useful to create regularly spaced data

Formulas and Equations I

- Numerical operators:

Task	Operator	Example	Result
Multiplication	*	2*3	6
Division	/	4/2	2
Exponent	^	2^3	8
Order of Operations	(..)	2*3+5 or 2*(3+5)	11 or 16
Power of ten	e or E	3.2e+2 or 3.2e-2	320 or 0.032

CQ: 10e4 in computer notation equals:

- A. 1,000
- B. 10,000
- C. 100,000
- D. Neither
- E. I don't know

Formulas and Equations II

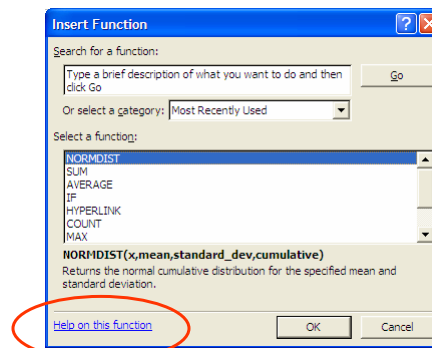
- Enter a formula which is calculated based on other cells
- Drag or Copy / Paste
 - Note that the result is different if you use absolute or relative references

	A	B	C	D
1	1	0	5	
2	2	0.1	5.2	
3	3	0.2	5.4	
4	4	0.3	5.6	
5	5	0.4	5.8	
6	6	0.5	6	
7	7	0.6	6.2	
8	8	0.7	6.4	
9	9	0.8	6.6	
10	10	0.9	6.8	
11		1	7	
12		1.1	7.2	
13		1.2	7.4	
14		1.3	7.6	
15		1.4	7.8	
16		1.5	8	
17				
18				

12

Pre-Programmed Functions

- Excel has lots of pre-programmed functions
 - E.g. normal distribution
 - Click f_x symbol to get a menu
 - Also look up the help files



13

Useful Pre-Programmed Functions

- AVERAGE()
- STDDEV()
- STDEV()
- MEDIAN()
- MAX()
- MIN()
- VAR()
- NORMDIST()
- NORMDISTINV()
- TDIST()
- TTEST()

CQ: I have used

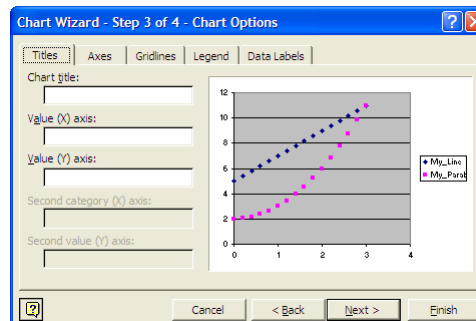
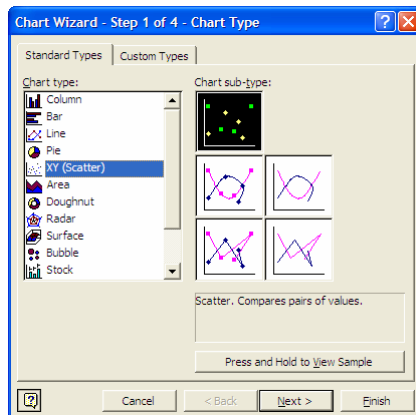
- A. all of these
- B. most of these
- C. a few of these
- D. none of these
- E. what are these?

These are only a few of the statistical functions, there are lots more!

14

Plotting in Excel

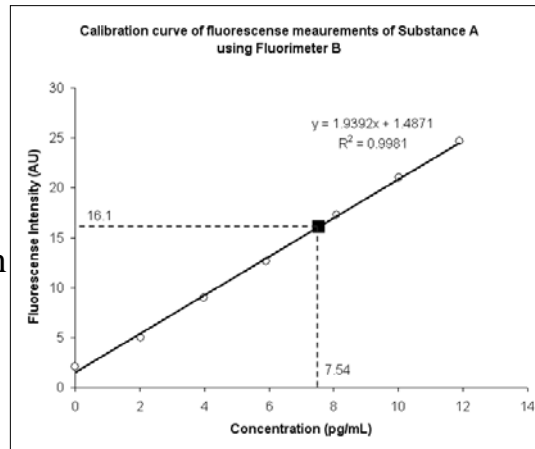
- Select data
- Choose Insert -> Chart



15

Proper Graph Formatting for Reports

- Label axes w/ Units (AU or arb. Units if needed)
- Independent variable on X-axis
- Dependent variable on Y-axis.
- Scatter (not line) plot
- Add a regression line (if appropriate)
- Descriptive title
- *Remember Sigfigs!*

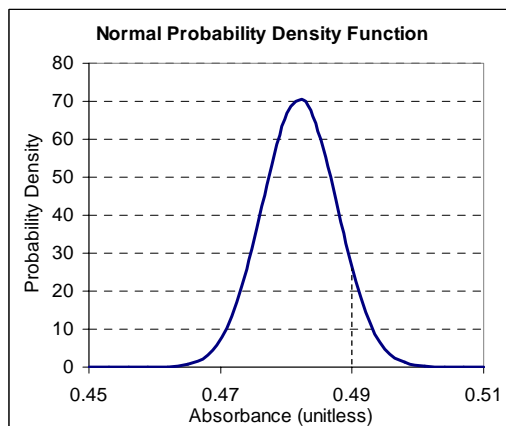


- Do not plot too many data sets on single graph - make multiple graphs instead

16

The Normal Error Curve

- Random errors are often distributed according to the normal error law



CQ: the probability that the absorbance is exactly 0.49 is:

- A. 0.49
- B. 25
- C. infinity
- D. zero
- E. I don't know

17

The Normal Error Law I

- The fraction of a population of observations whose values are between x and $x+dx$ is:

$$\frac{dN}{N} = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} dx$$

- I.e. the probability that an observation is between x and $x+dx$ is:

$$P(x, x+dx) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} dx$$

- *Probability density:*

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

18

The Normal Error Law II

- Cumulative probability
 - The probability that x has a value between x_1 and x_2 is:

$$P(x_1, x_2) = \int_{x_1}^{x_2} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} dx$$

- Normalized distribution:

$$z = \frac{x - \mu}{\sigma}$$

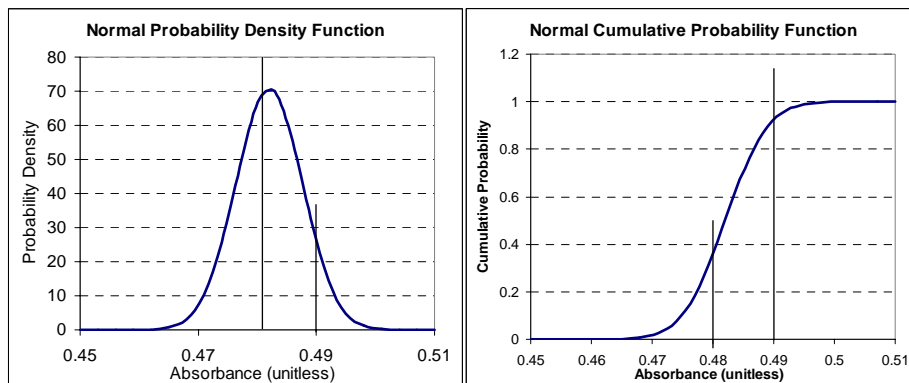
- In Excel

- Can type the whole formula (prone to errors)
- `NORMDIST(x, μ, σ, FALSE)` for PDF
 - E.g. `=NORMDIST(3,1,0.23,FALSE)`
- `NORMDIST(x, μ, σ, TRUE)` for CDF

19

PDF vs CDF

- Probability Density vs. Cumulative Probability



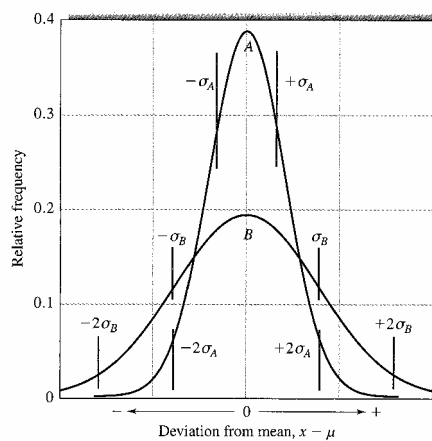
CQ: the probability that the absorbance is between 0.48 and 0.49 is:

- A. zero
- B. It is not defined
- C. 0.37
- D. 0.56
- E. I don't know

20

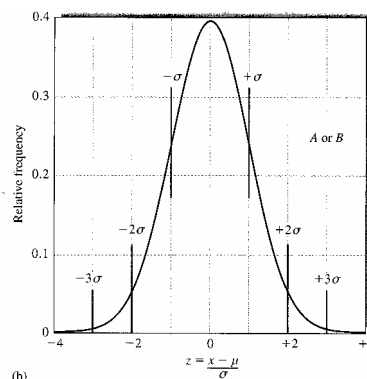
Normal Error Curves

With units:



(a)

Normalized:

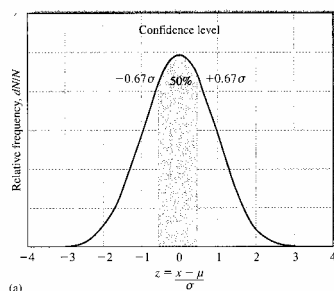


(b)

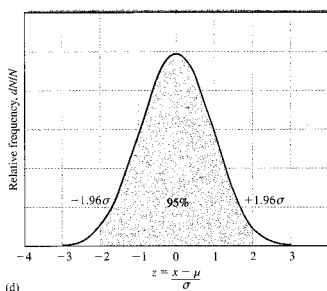
FIGURE a1-3 Normal error curves. The standard deviation of B is twice that of A; that is, $\sigma_B = 2\sigma_A$. (a) The abscissa is the deviation from the mean in the units of the measurement. (b) The abscissa is the deviation from the mean relative to σ . Thus, A and B produce identical curves when the abscissa is $z = (x - \mu)/\sigma$.

21

Areas under Gaussian Curve

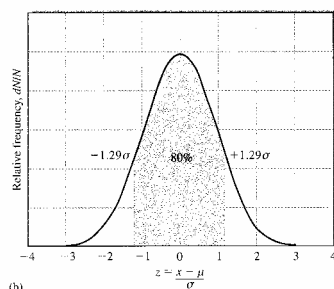


(a)

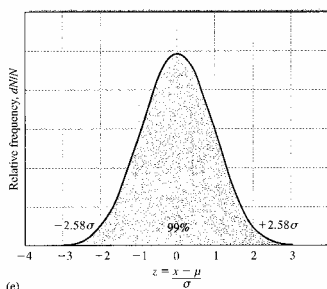


(d)

Also know that $\pm 1\sigma$ has 67%



(b)



(e)

Q: calculation w/ Excel?

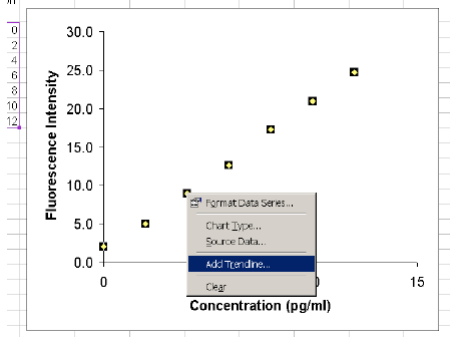
Linear Regression in Excel I

- Easy way (in graph)
- More complex way & more information (Analysis ToolPak)
- Example: calibration curve for fluorescence
- Input data from table:

Fluorescence Intensities	Concentration (pg/ml)
2.1	0
5.0	2
9.0	4
12.6	6
17.3	8
21.0	10
24.7	12

Linear Regression in Excel II

- Right-Click on data point:



Steps 2 & 3:

