# STATISTICAL ANALYSIS SUMMARY

## I. OBJECTIVE

To model the behavior of an experimental data set comprising a limited number of measurements using mathematical theory developed for an infinity of measurements and small, randomly distributed fluctuations (Gaussian distribution).

We would like to know:

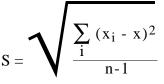
- How precise is our data
- How much confidence we have in the data
- How the error in the experimental data is propagated in calculated values and what are the main contributors to the propagated error
- How two experimental data sets compare to each other
- How an experimental value compares to an accepted literature value

 $x = \frac{\sum_{i} x_{i}}{r}$ 

### **II. MEASUREMENTS**

\* <u>Mean</u> of an experimental set:

\* Standard deviation of an experimental set:



\* What can be concluded from the mean and standard deviation is that:

- There is a 68% chance that a measurement will fall betwen  $x \pm S$
- There is a 95% chance that a measurement will fall between  $x \pm 1.96S$

# **III. CONFIDENCE INTERVAL**

\* The confidence interval  $\lambda$  is calculated from S, n and t. The latter is obtained from tables by setting P = 5 and DF = n - 1.

$$\lambda = \frac{t.S}{\sqrt{n}}$$

- \* The more measurements are made (i.e. the greater n or DF) the smaller the value of  $\lambda$
- \* What can be concluded from the confidence interval is:
  - I am 95% confident that the mean is within  $\pm \lambda$  of the "true" value (i.e. that which would be obtained if an infinity of measurements were performed)

$$x = x \pm \lambda$$
 (95% confidence, n = ...)

\* Special Cases:

• If measurements cannot be performed, the confidence interval may be estimated:

<u>x = x  $\pm \lambda$  (95% confidence, estimated)</u>

• Linear fitting of curves

- Reduce to linear expressions using mathematical manipulations

- Plot the data and model as y = m.x + b, where m and b are obtained with calculator, plotting software, or by applying:

$$m = \frac{n \cdot \sum_{i} (x_{i}y_{i}) - (\sum_{i} x_{i}) \cdot (\sum_{i} y_{i})}{n(\sum_{i} x_{i}^{2}) - (\sum_{i} x_{i})^{2}}$$
$$b = \frac{(\sum_{i} x_{i}^{2}) \cdot (\sum_{i} y_{i}) - (\sum_{i} x_{i}) \cdot (\sum_{i} x_{i}y_{i})}{n(\sum_{i} x_{i}^{2}) - (\sum_{i} x_{i})^{2}}$$

- For each data set, calculate the fluctuation with the model  $\delta y_i$  =  $y_i$  -  $(mx_i+b)$ 

- Calculate the standard deviations as necessary:

$$S_{y} = \sqrt{\frac{\sum_{i} (\delta y_{i})^{2}}{n-2}}$$

$$S_{m} = S_{y} \sqrt{\frac{n}{n(\sum_{i} x_{i}^{2}) - (\sum_{i} x_{i})^{2}}}$$

$$S_{b} = S_{y} \sqrt{\frac{(\sum_{i} x_{i}^{2})}{n(\sum_{i} x_{i}^{2}) - (\sum_{i} x_{i})^{2}}}$$

- Calculate  $\lambda_y$ ,  $\lambda_m$ , and  $\lambda_b$  by using the table with P = 5, DF = n-2 and  $\lambda = \frac{t.S}{\sqrt{n}}$ 

 $x = x \pm \lambda$  (95% confidence, best fit)

## **IV. SIGNIFICANCE TESTING**

\* Significance testing gives a measure of the accuracy of the experimental data

• Comparison with a literature value, μ:

- Calculate 
$$t = \frac{|x - \mu|}{S/\sqrt{n}}$$

- Knowing the DF (usually n-1), get P from the table.
- <u>There is a P% chance that the deviation from μ results from random fluctuations</u> and a (1-P)% chance that it is due to systematic error

• Comparison between two experimental data sets:

Calculate 
$$t = \frac{x_1 - x_2}{\sqrt{\frac{(n_1 - 1).S_1^2 + (n_2 - 1).S_2^2}{n_1 + n_2 - 2}}} \sqrt{\frac{(n_1 - 1).S_2^2}{n_1 + n_2 - 2}}$$

- Knowing  $DF = n_1 + n_2 - 2$ , obtain P from the table

- <u>There is a P% chance that the deviation between x<sub>1</sub> and x<sub>2</sub> is the result of random fluctuations and a (1-P)% chance that it is due to systematic error</u>

### **V. ERROR PROPAGATION**

- \* Verify that the formula only involves independent variables
- \* Decide which variables are known with certainty to simplify the calculations
- \* Evaluate all others  $\lambda_i$
- \* Calculate the average value using the means: A = f(x,y,z)
- \* Calculate the partial derivatives of the function with respect to each variable, holding all other variables constant, and evaluate the partial derivatives at x, y and z
- \* Calculate  $\lambda_A$  from:

$$\lambda_{A} = \sqrt{\left(\left(\frac{A}{x}\right)^{2}_{y,z} \cdot \lambda_{x}^{2} + \left(\frac{A}{y}\right)^{2}_{x,z} \cdot \lambda_{y}^{2} + \left(\frac{A}{z}\right)^{2}_{x,y} \cdot \lambda_{z}^{2}\right)}$$

<u>A = A  $\pm \lambda_A$  (95% confidence, propagated)</u>