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

II. MEASUREMENTS

* Mean of an experimental set: $x = \frac{\sum_i x_i}{n}$

* Standard deviation of an experimental set: $S = \sqrt{\frac{\sum_i (x_i - x)^2}{n-1}}$

* 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 betwen $x \pm 1.96S$

III. CONFIDENCE INTERVAL

* The confidence interval λ 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 λ

* 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:

$$\underline{x = x \pm \lambda \text{ (95\% confidence, estimated)}}$$

- 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:

$$m = \frac{n.\sum_i (x_i y_i) - (\sum_i x_i).(\sum_i y_i)}{n(\sum_i x_i^2) - (\sum_i x_i)^2}$$

$$b = \frac{(\sum_i x_i^2).(\sum_i y_i) - (\sum_i x_i).(\sum_i x_i y_i)}{n(\sum_i x_i^2) - (\sum_i x_i)^2}$$

  - 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 λ_y, λ_m, and λ_b by using the table with P = 5, DF = n-2 and $\lambda = \frac{t.S}{\sqrt{n}}$

$\underline{x = x \pm \lambda}$ (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.

- There is a P% chance that the deviation from μ results from random fluctuations and a (1-P)% chance that it is due to systematic error

• Comparison between two experimental data sets:

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

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

- There is a P% chance that the deviation between x_1 and x_2 is the result of random fluctuations and a (1-P)% chance that it is due to systematic error

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 λ_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 λ_A from:

$$\lambda_A = \sqrt{\left(\left(\frac{A}{x}\right)^2_{y,z}.\lambda_x^2 + \left(\frac{A}{y}\right)^2_{x,z}.\lambda_y^2 + \left(\frac{A}{z}\right)^2_{x,y}.\lambda_z^2\right)}$$

$\underline{A = A \pm \lambda_A}$ (95% confidence, propagated)