

Measurement and Experimental Error

References: 1) An Introduction to Error Analysis, 2nd Edition, J. R. Taylor
2) Data Reduction and Error Analysis for the Physical Sciences, 2nd Edition, P. R. Bevington, D. K. Robinson

1.1 Definitions

- Error analysis is the evaluation of uncertainty in measurements
- Error means uncertainty, not mistake, no negative connotation.
- Experimentation => measurements => comparison to a standard
e.g., length, time, weight, volume, chemical composition etc.

- All measurements have associated “error” or uncertainty.
- Experimental results should be accompanied with best estimate of the quantity and range within which you are confident e.g., $x \pm \Delta x$
 x = best estimate (want this to be as exact as possible)
 Δx = uncertainty (this is really an estimate)
relative uncertainty: $\Delta x/x$
- **Precision:** measure of how carefully the result is determined without reference to any true value (precision is related to uncertainty, e.g., Δx)
- **Accuracy:** how close the measurement is to the exact value

Example 1.

Measurement of length.

Questions:

- Why would repeated measurements give different numbers (what is the cause)
- What is the “true” length. (What actually do we mean by the length of the object; there is no such thing, only a range of lengths).
- How do I report the length?
- How can I minimize the “error” or uncertainty in the measurement?

- One goal of an experimentalist is to minimize the uncertainties in a measurement. You should also always state how your uncertainties were calculated.

Example 2.

Measurement of particle number concentration with a CNC or CPC.

Concentration = No. particles/volume of air

- Numerator: counts (C), Poisson uncertainty $\{\Delta C = \sqrt{C}\}$
- Denominator, volume of air sampled = volume flow rate*time, each has an uncertainty (e.g., ΔQ and Δt).

1.2 How To Evaluate Uncertainties

- Reading scales (ruler, graduated cylinder) ~ 1/2 smallest graduation
- Use manufacturers stated precision

- Use \pm smallest digit on digital readout (i.e., stop watch) {this is really not good because reaction time is greatest cause of uncertainty}.
- In situations of counting, relative uncertainty is $1/\sqrt{\text{counts}}$
- Repeat the measurement and use the range, or better yet, the standard deviation – really we need a statistical analysis; (typically, the better the analysis the lower the uncertainty)
- (state what you do, i.e., how you estimated the uncertainty)

In some cases there is no way to compare the measurement to a correct value, (i.e., measurements of particle chemical composition). Normally, compare your measurement to a more accurate standard. If you have no accurate standard, you may have a very precise measurement but it could be far off from the correct answer. Need a “gold standard” to test accuracy of measurement.

1.3. Systematic/Random Errors

Systematic: same magnitude and sign under same conditions, e.g., suppose device to measure flow rate was always systematically low, all else in the measurement is correct. (If plot our CN to “true” CN linear regression will give 1:1 slope but there will be a non-zero intercept).

Types of systematic errors.

- 1) Natural; i.e., thermal expansion of materials, effect of moisture
- 2) Instrumental; Instrument imperfections, i.e., made wrong or damaged, (stretched ruler)
- 3) Personal, due to habits of observer (i.e., improper use of a pipet)

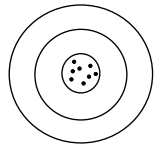
**** can't use statistical methods to analyze**

Random:

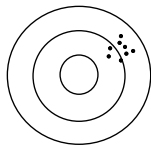
- Variable magnitude and sign.
- Due to fact that it is impossible to exactly duplicate any measurement.
- Minimized by repeating the measurement and taking the mean.

An Example: SHOOTING EXPERIMENT

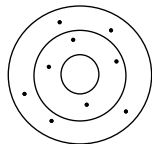
- Random errors (PRECISION) can be estimated from repeated experiments.
- Systematic errors (ACURRACY) require a gold standard for comparison.



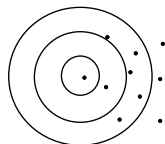
random: small
Systematic: small



random: small
Systematic: large



random: large
Systematic: small



random: large
Systematic: large

error = radial distance to center (or mean)

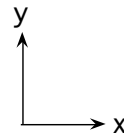
No "gold" standard



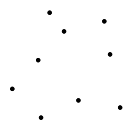
random: small
Systematic: small



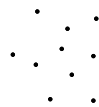
random: small
Systematic: large



want as small random error as possible, can't say much about the bottom situation



random: large
Systematic: small



random: large
Systematic: large

What is the correct answer?

Can't assess systematic error if no gold standard (accuracy)

Random error (precision) can be determined

1.4 Accuracy and Significant Figures

- Accuracy: how close the measurement is to the exact value
- Perfect accuracy requires an infinite number of digits.
- The number of correct digits in a measurement indicates the accuracy of the measurement. The accuracy is implied from the number of **significant figures** reported.
- How many significant figs to carry when:
 - adding or subtracting: round off answer to number of sig. figs. of least factor.
 - multiplying or dividing; keep same number of significant figs as number with least sig. figs. (i.e., $236.52/1.75 = 135$ not 135.1443)
 - (Note: there is a difference between +- or x / operations, the latter produces more significant figures, unless integers)

Rounding off: how to drop excess digits,

- if digits dropped are less than 5, 50, ... Make no change, i.e., 3.234, 2 sig figs is 3.2.
- if digits dropped are greater than 5, 50, ... Last retained digit is increased, i.e., 3.256 is 3.3
- if digit dropped is 5, round off preceding digit to an even number, i.e., 3.25 is 3.2, 3.35 is 3.4 etc.

Note, use all significant figures in the calculations, round off answer to appropriate sig. figs.

1.5 Reporting Accuracy of Measurements

Use an associated uncertainty to indicate the accuracy of the measurement

- **Absolute** uncertainty is Δx , ie $CN = 590. \pm 18 \text{ cm}^{-3}$
- **Relative** uncertainty, is $\Delta x/x$, ie. $CN = 590 \text{ cm}^{-3} \pm 3\%$ (also referred to as the **precision**, i.e., Δx of 1 inch in 1 mile is a precise measurement, Δx of 1 inch in 5 inches is an imprecise measurement.

When reporting measurements uncertainties, typical one significant fig. is best, 2 sig figs in cases of high precision.(unless the first digit is a 1 or 2, i.e., don't write ± 0.1 do write ± 0.12)

The reported value (i.e., x in $x \pm \Delta x$) last sig fig should be of similar order of magnitude as the uncertainty, i.e., 56.01 ± 3 should be 56 ± 3 , the uncertainty determines how you write the best estimate value.

1.6 Propagation of Errors.

Generally, determining a physical quantity requires 1) making a measurement, 2) determining the value using some expression with other measured variables. Thus, we need a way to calculate uncertainty of final answer.

Example, CN counter; I measure a count of 9833 particles per 1 sec sampling at 1 l/min. Say my timer is accurate to ± 0.001 sec and my flow rate is accurate to within 2%

($\Delta Q = 0.02$ L/min). What is the accuracy I should report with my measured concentration.

Method 1: Exact Solution; it can be shown that the final error due to the combined **independent and random** errors of individual measurements is (with negligible approximation, 1st order Taylor series expanded about 0, -see Taylor Chapt 3 for detailed derivations):

$$\Delta y = |\Delta x_1 df/dx_1| + |\Delta x_2 df/dx_2| + |\Delta x_3 df/dx_3| + |\Delta x_4 df/dx_4| + \dots \text{ (partial derivatives),}$$

If the errors are independent, then can expect some cancellation between errors and the total error will be less. Thus, better to use quadrature sum in above equation.

For our example; $CN = C/(Q t)$: $\Delta CN = |\Delta C 1/(Qt)| + |\Delta Q Ct/(Qt)^2| + |\Delta t CQ/(Qt)^2|$

$$\Delta CN = \text{sqr}(9833)/(16.67 \text{ cm}^3/\text{s} * 1 \text{ s}) + (0.33 \text{ cm}^3/\text{s})(9833)(1\text{s})/(16.67 \text{ cm}^3/\text{s} * 1 \text{ s})^2 + (0.001\text{s})(9833)(16.67 \text{ cm}^3/\text{s}) / (16.67 \text{ cm}^3/\text{s} * 1 \text{ s})^2$$

$$\Delta CN = 5.95 + 11.68 + 0.59 [1/\text{cm}^3] = 18.22 1/\text{cm}^3 \text{ and the concentration is } 590 \pm 18$$

Using quadrature sum: $\Delta CN = \text{sqr}[5.95^2 + 11.68^2 + 0.59^2] = 13.12$ (less than above).

Method 2. A simpler approach is to use propagation of relative errors,

Operations involving multiplying and dividing the overall uncertainty can be calculated from the square root of the sum of the squares of **relative** errors. (Note, for operations involving adding or subtracting overall uncertainty can be calculated from the square root of the sum of the squares of **absolute** errors).

Our example; $(\Delta CN/CN)^2 = (\Delta C/C)^2 + (\Delta Q/Q)^2 + (\Delta t/t)^2$
 $(\Delta CN/CN)^2 = (0.01)^2 + (0.02)^2 + (0.001)^2 = 5.03 \text{ E- } 4$ or $\Delta CN/CN = 0.022$ and the answer is 590 ± 13 .

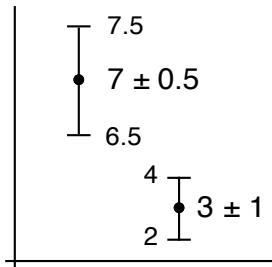
Method 3. Use the range in the final answer as the uncertainty.

1.7 Comparing Different Measurements or Measurements to Models

Compare error bars, if overlap than there is no difference between measurements or no discrepancy. (Discrepancy is the difference in the estimated quantity). If the discrepancy is larger than the combined margin of error than the numbers do not agree). (Combined margin of error is sum of absolute uncertainties: add uncertainties for + or – operations)

Numbers compared could be: mean \pm std error; slope $\pm 1\sigma$...

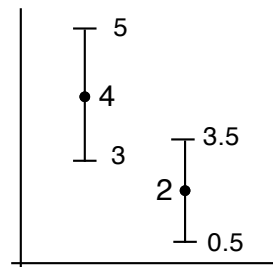
Comparing numbers: Is there a difference?



Difference: $(7-3) \pm (0.5 + 1)$

4 ± 1.5

Yes, there is a difference



Difference: $(4-2) \pm (1 + 1.5)$

2 ± 2.5

No, there is a difference
(or can't tell from these data)

1.8 Statistical Analysis – Linear Regression

- mean and standard deviation can be calculated from the data
- a statistical analysis can be used to make more detailed calculations (descriptions of the data) if confident in the parent distribution that describes the spread of the data points.
- The most common distributions that describes experimental results:

Gaussian: represents random processes

Poisson: probability of observing a specific number of counts N ($\Delta N = \sqrt{N}$)

Repeated experiments: If you repeat an experiment you can use the (random) variation in the measurements to estimate uncertainties due to random errors.

e.g., **error of the mean** = std/\sqrt{N}

The error of the mean, (or standard error), is the uncertainty associated with the mean. For similar expts. the uncertainty of each measurement is similar (\sim stdev), thus to reduce

uncertainty of mean, repeat the experiment (however, standard error is weak function of N)

Mean \pm standard error	multiple measurements
$X \pm \text{stdev}$	one measurement

Confidence intervals:

1- σ = 68.3% confidence interval: Eg: $x \pm \Delta x$, where $\Delta x = \sigma = \text{stdev}$: this means that you are 68% confident that x is within $x-\Delta x$ and $x+\Delta x$, if random errors are normally distributed (or there is a 68% probability the event will be within $\pm 1 \sigma$ either side of the measurement (area under normal curve ± 1 stdev from mean is 68%).

2- σ = 95.4% confidence intervals:

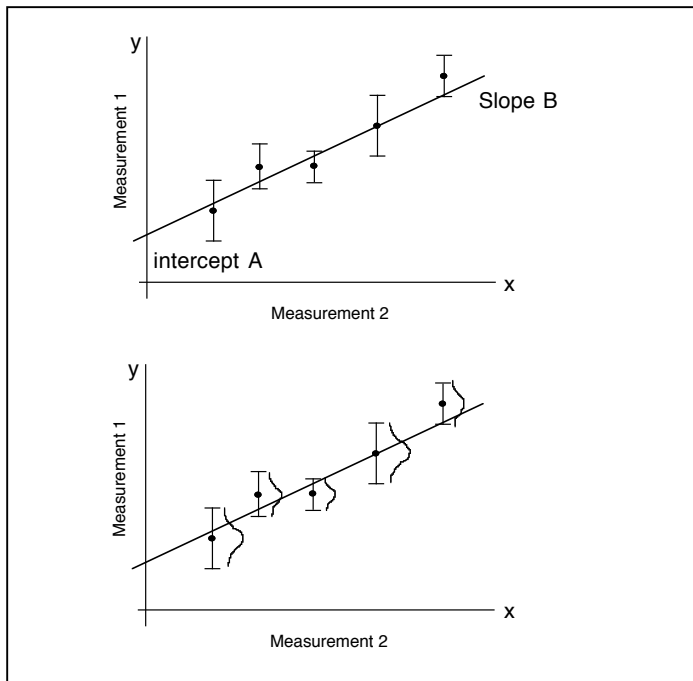
3- σ = 99.7% confidence intervals:

Least Squares Curve Fitting

Linear Regression:

Testing data that should fit a straight line (can do least squares fit on any function)

- i.e., comparing measurements of the same thing (instrument intercomparison, eg slope is how well the measurements agree, intercept is evidence for systematic differences)
- e.g., comparing measurements to theory.



Goal of linear regression is to find A and B.

Assumes

- no uncertainty in x
- y uncertainties are all the same magnitude (if no weightings are used)
- each y -measurement is governed by normal dist. and with same σ_y .

Linear Eq. $y_i = A + Bx_i$

Best estimate of A and B is values for which χ^2 is a minimum (\Rightarrow least squares).

$$\chi^2 = \sum_{i=1}^{\text{all data points}} \frac{(y_i - A - Bx_i)^2}{\sigma_y^2}$$

(σ_y : width of normal distribution that determines variability in the measurements)

Results from a Linear Regression

Coefficient of Linear Correlation (r)

- Pearson product moment coefficient of correlation, or correlation coefficient
- Measures strength of linear relationship between two variables (x & y). or how well the points fit a straight line.
- $r \leq \pm 1$
 - $r = 1$ positive linear correlation
 - $r = -1$ negative linear correlation
 - $r = 0$ no correlation

Coefficient of Determination (or of variation), r^2

- indicates percent of variability in the y value explained by x variable. $r^2 \leq 1$

Uncertainties in the slope and intercept and Hypothesis testing

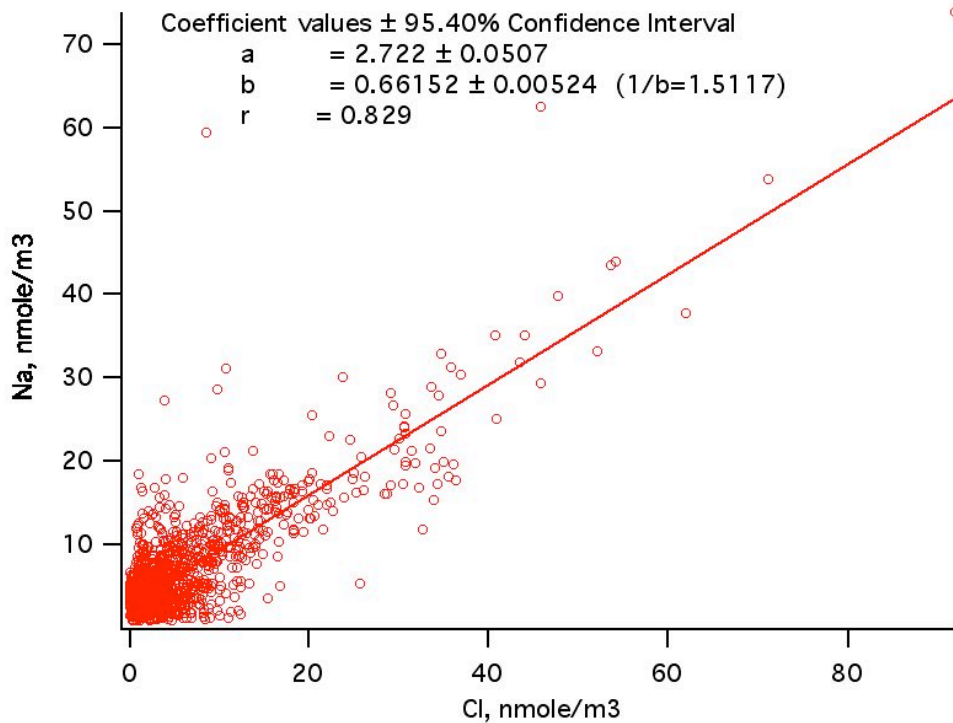
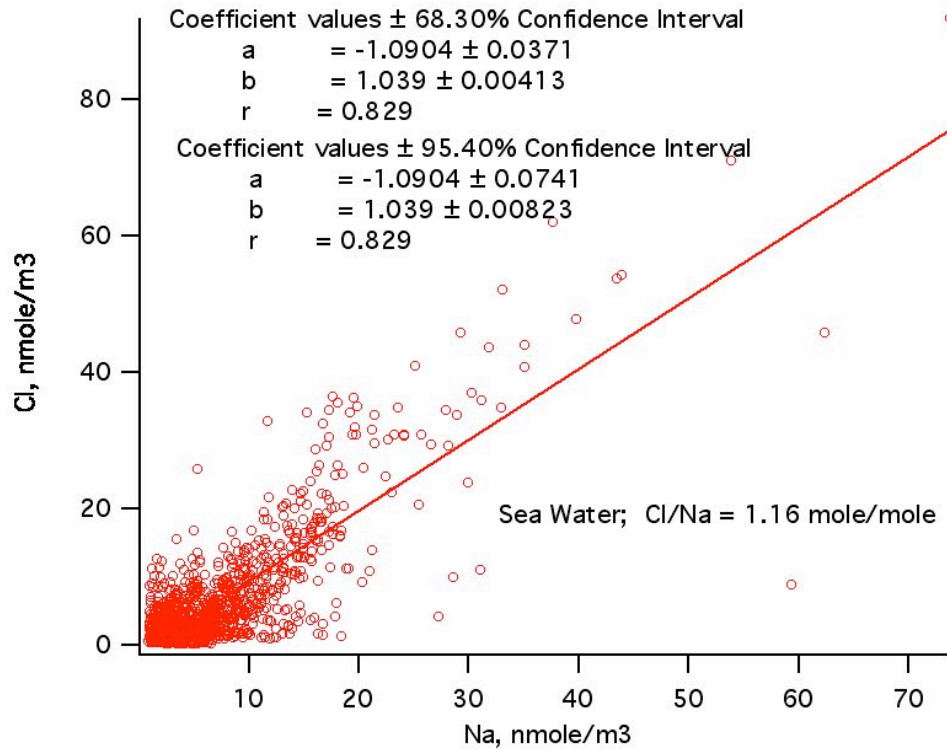
- use commercial software to calculate uncertainties associated with the slope and intercept, i.e., IGOR, can choose the uncertainty at desired confidence interval, or software may give the standard deviation associated with the slope and intercept – you can then calculate the desired confidence interval.
- Typical hypothesis to test is that the slope is 0 (i.e., x isn't useful in linearly predicting y). e.g., given the "t-ratio" and degrees of freedom, one can then test hypothesis at desired α -level (ie, $\alpha = 0.05$ is common, area under one end of t-curve is $0.05/2$). If determined t (from table) is less than slope's t-ratio, reject hypothesis and accept null hypothesis, slope $\neq 0$. The corresponding confidence interval is determined by $\sigma \times t$ (from table)

Deming Least Squares

What if assumption 1) is not true, x variable also has uncertainty.

- The least squares calculation must minimize both x and y errors.
- This is the situation for comparing measurements and is needed when calculating slope and intercepts
- Solution: use a Deming Least Squares Fit.

Data from ACE-Asia, C-130 PILS-IC (n= 28,862)



1.9 Calibration Curves: How to measure a quantity of interest.

Analytical equipment is usually optimized to detect and report chemical information of one specific factor (chemical reaction, concentration, analyte, or component of the system under investigation) within a specific range of parameters and factor levels. However, not all levels of the factor can be detected. In order to accurately describe the instrument response to factor level intensity, it is necessary to test the response against a series of known standards. This is called calibrating the instrument. The result will be a plot of response vs. factor intensity, known as the **calibration curve**.

Calibration curves are functions of an instrument's responses to a range of **factor levels**. A factor level could be concentration, or chemical potential, or some other measurable quantity. In some cases, the factor level may be too low to be detected. In some cases, the factor level may be too high and overload the detector. The optimal range of a factor level for detection by an instrument is usually narrow. Ideally, this narrow range represents a linear response or output proportional to the factor level intensity. If the factor level increases, the response increases. If the factor level decreases, the response decreases.

The calibration curve has three distinct regions. (See figure 1). The segment **Below the Limit of Detection** is where the factor level is not great enough or intense enough to produce a signal greater than the background noise of the instrument. There is a high **noise to signal ratio**. The **Linear Region** is the region where the response is proportional to the factor level. As factor level increases, response increases. The linear region extends to the **Limit of Linearity**, or the maximum factor level that will produce a linear response. Any measurements of factor levels greater than the limit of linearity will fall in the region **Beyond the Limit of Detection**, or where the detection device becomes overloaded with incoming data.

The integrity of a calibration curve depends on how much of the entire linear range is defined. In figure 2, three points denote a linear portion of the calibration curve, but there is great uncertainty. Are these points merely coincidence or are they in the linear region? In order to accurately describe such a curve, a series of known standards must be constructed to add detail to the wide interval between points (figure 3)

Figure 1: Calibration Curve

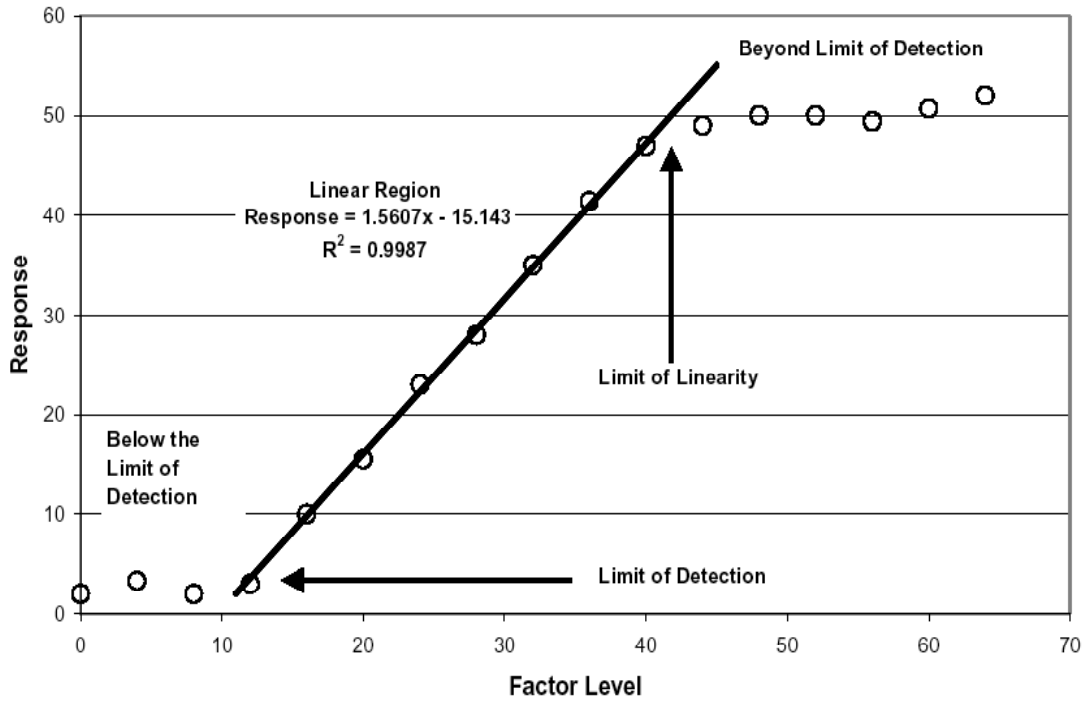


Figure 2: Calibration Curve

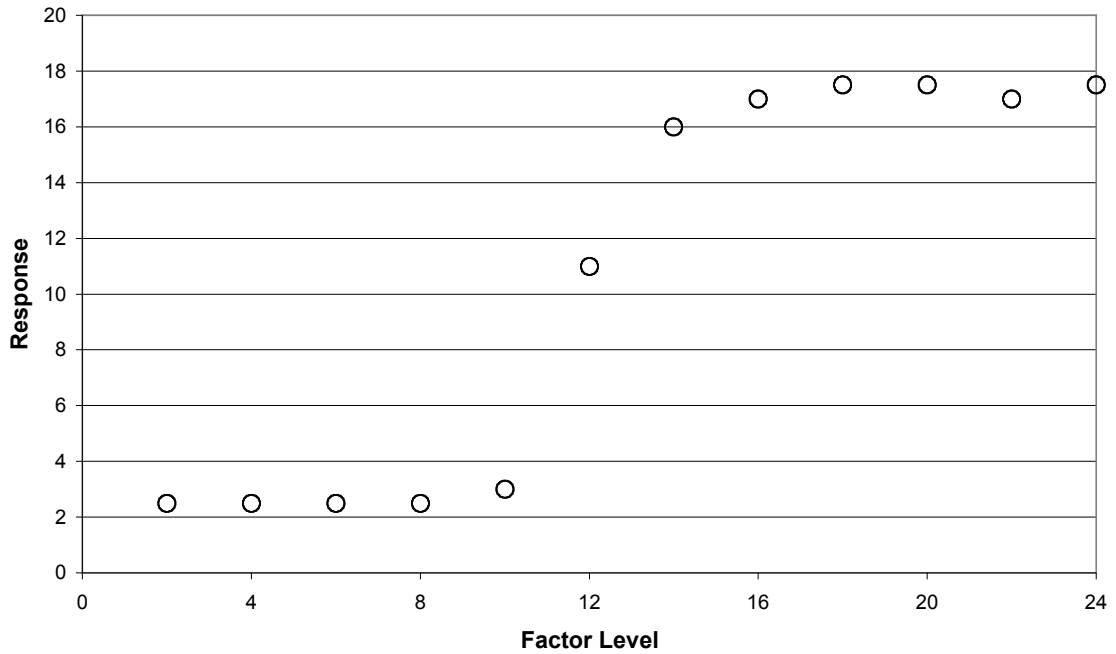
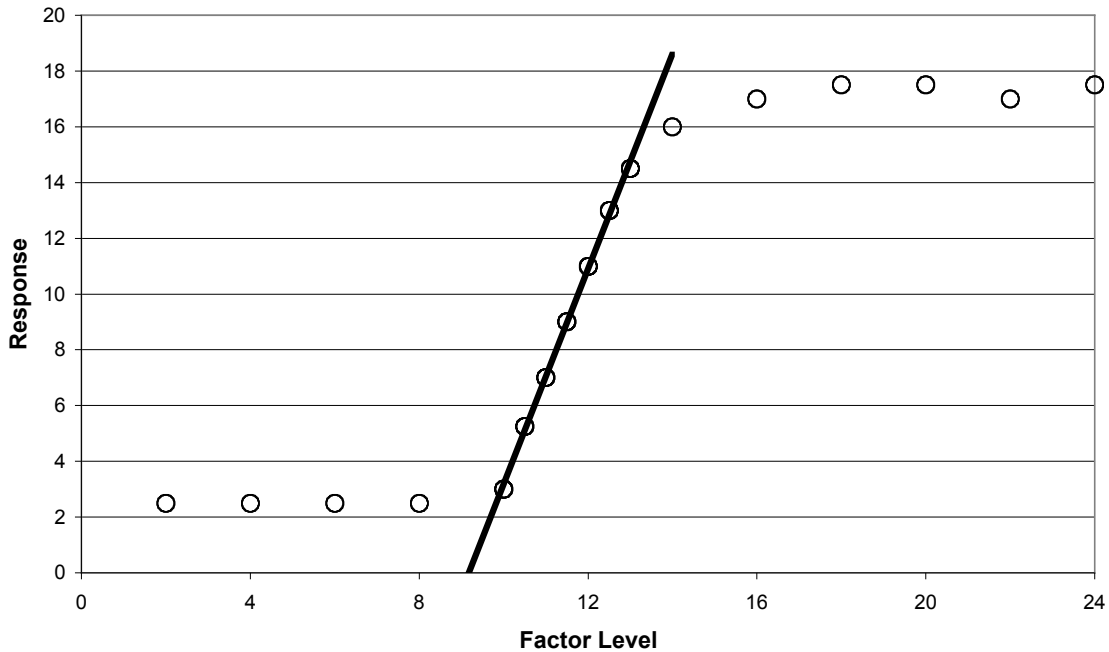


Figure 3: Calibration Curve



Once a reliable calibration curve has been determined, factor levels for unknown samples can be determined. The general formula for the linear portion of a calibration curve is

$$response = (slope \times FL) + b$$

where, FL is the factor level (intensity, concentration, i.e., unknown to be determined, or the input) and response is the instrument's output. The slope is the analytical slope for the detection of that factor. b is the Y-intercept (note, in many cases $b = 0$). Once the calibration curve is constructed from a series of standards, the unknown concentrations of a factor in a solution can be determined by rearranging the above equation to the following.

$$FL_{unk} = x_j = \frac{response - b}{slope}$$

Of course, for the calibration curve to be effective the factor must be detected and a response provided. If the response of an unknown falls outside the established linear range of the calibration curve, then additional work is required. If the response falls below the Limit of Detection, then one must concentrate the sample. If the response lies above the Limit of Linearity, then one must dilute the sample.

NOTE: All measurements (data points) and calibration curves have statistical limitations. The following pages contain information and mathematical equations that describe some of the basic statistics required for quantitative work for this course. Other possible references include: **Quantitative Chemical Analysis**, **Chemometrics**, or **Instrumental Analysis** textbook for a complete inventory of statistics for chemistry. Many statistics texts will also be helpful.