Chapter 3: Data Collection and Preliminary Data Analysis

3.1 Generalized measurement system
3.2 Performance characteristics of sensors and sensing systems
3.3 Data validation and preparation
3.4 Descriptive measures of sample data
3.5 Plotting data
3.6 Overall measurement uncertainty
3.7 Propagation of errors
3.8 Planning a non-intrusive field experiment

3.1 Generalized Measurement System

- Three stages:
  - Detector-transducer
  - Intermediate
  - Output or terminating
- Unfortunately spurious inputs corrupt readings:
  - Interfering inputs (direct impact)- such as solar radiation on thermocouple measurement
  - Modifying inputs (more subtle)

Fig. 3.1 Schematic of the generalized measurement system

Fig. 3.2 Different types of inputs and noise in a measurement system
3.2a Performance Characteristics of Sensors and Sensing Systems

Fig. 3.4 Concepts of accuracy and precision illustrated in terms of shooting at a target

(c) Span or dynamic range- range of variation (min to max)

3.2b Performance Characteristics of Sensors and Sensing Systems

Fig. 3.5 Concepts of threshold and resolution

Fig. 3.6 Zero drift and sensitivity drift

(recall from Chap1 that static sensitivity K is the slope of the response)
3.2c Performance Characteristics of Sensors and Sensing Systems

Output signal

Maximum hysteresis \( \pm \frac{cd}{2} \)

Local hysteresis \( \pm \frac{ab}{2} \)

Input signal

Fig 3.7 Illustrative plot of a hysteresis band of a sensor showing local and maximum values

3.2d Performance Characteristics of Sensors and Sensing Systems

Calibration involves determining a relationship between instrument raw reading and the best estimate of the true value of the quantity being measured using a more accurate primary instrument. Usually bias only is corrected.

Fig. 3.8 Static calibration to define bias and random variation or uncertainty. Note that \( s \) is the standard deviation of the deviations between measurement and the least squares model (from Doebelin 1995 by permission of McGraw-Hill)
3.2e Performance Characteristics of Sensors and Sensing Systems

Fig. 3.9 Concept of rise time of the output response to a step input

Recall prior concepts of “time constant” and settling time

Fig. 3.10 Effects of frequency response and phase-shift response on complex waveforms (from Holman and Gajda 1984 by permission of McGraw-Hill)

3.2.2 Types and Categories of Measurements

A **primary measurement**: one that is obtained directly from the measurement sensor. The basic criterion is that a primary measurement is of a single item from a specific measurement device.

A **derived measurement** is one that is calculated using one or more measurements either primary or derived

Further, measurements can also be categorized by type:

- **Stationary data** does not change with time (mass of water in a tank, the area of a room, …)
- **Time dependent data** varies with time (pollutant concentration in a water stream, temperature of a space, the chilled water flow to a building. Further, one differentiates:
  - Time-series data consist of a multiplicity of data taken at a single point or location over fixed intervals of time, thus retaining the time sequence nature.
  - Cross-sectional data are data taken at single or multiple points at a single instant in time with time not being a variable in the process.
3.2.3 Data Recording Systems

(a) **Recording interval** is the time period or intervals at which data is recorded (a typical range for a thermal systems could be 1-15 min)

(b) **Scan rate** is the frequency with which the recording system samples individual measurements; this is often much smaller than the recording interval (with electronic loggers, a typical value could be one sample per second)

(c) **Scan interval** is the minimum interval between separate scans of the complete set of measurements which includes several sensors (a typical value could be 10-15 seconds)

(d) **Non-process data trigger**. Need to avoid measuring non-process data (i.e., temperature data when the flow in a pipe is stopped but the sensor keeps recording the temperature of the fluid at rest). Often threshold trigger used (for example, whether the pump which induces the flow thru the pipe is operational or not).

3.3a Data Validation and Preparation

Data Reduction- process of distilling raw data into a form that is suitable for subsequent analysis

(a) **averaging** involves removing gross or egregious errors

(b) **limit checks**:  
- physical limits: relative humidity <=100%  
- expected limits: indoor air relative humidity 30-60%  
- theoretical limits: efficiency of power plant < Carnot

(c) **Independent/ consistency checks** involving mass and energy balances (it is advisable to have redundancy)
3.3b Data Validation and Preparation

(d) Outlier rejection by visual means

Fig. 3.11 Scatter plot of the hourly chilled water consumption in a commercial building. Some of the obvious outlier points are circled (from Abbas and Haberl 1994 by permission of Haberl).

(e) Handling missing data

(a) Use observations with complete data only: simplest and most obvious, and is adopted in most analysis. Many of the software programs allow such cases to be handled. Instead of coding missing values as zero, analysts often use a default value such as -99 to indicate a missing value. This approach is best suited when the missing data fraction is small enough not to cause the analysis to become biased.

(b) Reject variables: In case only one or a few channels indicate high levels of missing data, the judicious approach is to drop these variables from the analysis itself. If these variables are known to be very influential, then more data needs to be collected with the measurement system modified to avoid such future occurrences.
3.3b Data Validation and Preparation

(e) Handling missing data (contd.)

(c) *Adopt an imputation method* or data rehabilitation (*could cause bias!*)

(i) Substituting by a constant value (may distort the probability distribution of the variable, its variance and its correlation with other variables);

(ii) Substituting by the mean of the missing variable deduced from the valid data (same problem but would perhaps add a little more realism to the analysis);

(iii) univariate interpolation (different methods are available- see numerical methods textbooks)

(iv) Use a regression model

---

3.4a Descriptive Measures of Sample Data

3.4.1 Summary statistical measures

a) **Mean** (or arithmetic mean or average) of a set or sample of $n$ numbers is:

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

where $n$ = sample size, and $x_i$ = individual reading

(b) **Weighted mean** of a set of $n$ numbers is:

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

where $w_i$ is the weight for group $i$.

(c) **Geometric mean** is more appropriate when studying phenomenon that exhibit exponential behavior (like population growth, biological processes, ...). This is defined as the $n$th root of the product of $n$ data points:

$$
X_{\text{geometric}} = \left[ x_1 \cdot x_2 \cdots x_n \right]^{1/n}
$$
3.4b Descriptive Measures of Sample Data

3.4.1 Summary statistical measures

(d) **Mode** is the value of the variate which occurs most frequently. When the variate is discrete, the mean may turn out to have a value that cannot actually be taken by the variate. In case of continuous variates, the mode is the value where the frequency density is highest. For example, a survey of the number of occupants in a car during the rush hour could yield a mean value of 1.6 which is not physically possible. In such cases, using a value of 2 (i.e., the mode) is more appropriate.

(e) **Median** is the middle value of the variates, i.e., half the numbers have numerical values below the median and half above. The mean is unduly influenced by extreme observations, and in such cases the median is a more robust indicator of the central tendency of the data. In case of an even number of observations, the mean of the middle two numbers is taken to be the median.

(f) **Range** is the difference between the largest and the smallest observation values.

(g) **Percentiles** are used to separate the data into bins. Let p be a number between 0 and 1. Then, the \((100p)\text{th percentile (also called pth quantile)}\), represents the data value where \(100p\%\) of the data values are lower. Thus, 90\% of the data will be below the 90\text{th percentile}, and the median is represented by the 50\text{th percentile}.

3.4c Descriptive Measures of Sample Data

3.4.1 Summary statistical measures

(h) **Inter-quartile range** (IQR) cuts out the more extreme values in a distribution. It is the range which covers the middle 50\% of the observations and is the difference between the lower quartile (i.e., the 25\text{th percentile}) and the upper quartile (i.e., the 75\text{th percentile}). In a similar manner, **deciles** divide the distribution into tenths, and **percentiles** into hundredths.

(i) **Deviation** of a number \(x_i\) in a set of \(n\) numbers is a measure of dispersion of the data from the mean, and is given by:

\[
d_i = (x_i - \bar{x})
\]

(ii) (j) The **mean deviation** of a set of \(n\) numbers is the mean of the absolute deviations:

\[
\bar{d} = \frac{1}{n} \sum_{i=1}^{n} |d_i|
\]
3.4d Descriptive Measures of Sample Data

3.4.1 Summary statistical measures

(k) The variance or the mean square error (MSE) of a set of n numbers is:

\[ s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{s_{xx}}{n-1} \]

where \( s_{xx} \) = sum of squares

(l) The standard deviation of a set of n numbers

\[ s_x = \left( \frac{x_{xx}}{n-1} \right)^{1/2} \]

The more variation there is in the data set, the bigger the standard deviation

(m) Coefficient of variation is a measure of the relative error, and is often more appropriate than the standard deviation. It is defined as the ratio of the standard deviation and the mean:

\[ CV = \frac{s_x}{\bar{x}} \]

This measure is also used in other disciplines: the reciprocal of the “signal to noise ratio” widely used in electrical engg, and also used as a measure of “risk” in financial decision making.

3.4e Descriptive Measures of Sample Data

(n) Trimmed mean. The sample mean may be very sensitive to outliers, and hence bias the analysis results.

The sample median is more robust since it is impervious to outliers.

However, non-parametric tests which use the median are less efficient than parametric tests in general.

Hence, a compromise is to use the trimmed mean value which is less sensitive to outliers than the mean but more sensitive than the median. One selects a trimming percentage \( 100r\% \) with the recommendation that \( 0 < r < 0.25 \).

Suppose one has a data set with \( n=20 \). Selecting \( r=0.1 \) implies that the trimming percentage is 10% (i.e., two observations). Then, two of the largest values and two of the smallest values of the data set are rejected prior to subsequent analysis.
### Example 3.4.1- Summary Statistics

<p>| | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2.97</td>
<td>3.00</td>
<td>3.20</td>
<td>3.44</td>
<td>3.94</td>
<td>4.08</td>
<td>4.28</td>
<td>4.46</td>
<td>4.76</td>
<td>5.04</td>
<td>5.32</td>
</tr>
<tr>
<td>5.60</td>
<td>5.85</td>
<td>6.04</td>
<td>6.13</td>
<td>6.23</td>
<td>6.32</td>
<td>6.52</td>
<td>6.72</td>
<td>6.82</td>
<td>6.98</td>
<td>7.16</td>
</tr>
<tr>
<td>7.62</td>
<td>7.71</td>
<td>7.81</td>
<td>7.93</td>
<td>8.11</td>
<td>8.29</td>
<td>8.47</td>
<td>8.54</td>
<td>8.62</td>
<td>8.70</td>
<td>8.79</td>
</tr>
</tbody>
</table>

#### Count 90

Average 10.0384

Median 9.835

Mode

Geometric mean 9.60826

5% Trimmed mean 9.98444

Variance 8.22537

Standard deviation 2.86799

Coeff. of variation 28.5701%

Minimum 2.97

Maximum 18.26

Range 15.29

Lower quartile 7.93

Upper quartile 12.16

Interquartile range 4.23

### 3.4f Descriptive Measures of Sample Data

#### 3.4.2 Covariance and Pearson correlation coefficient

Though a scatter plot of bivariate numerical data gives a good visual indication of how strongly variables x and y vary together, a quantitative measure is needed. This is provided by the covariance which represents the strength of the linear relationship between the two variables:

\[
\text{cov}(xy) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})
\]

where \( \bar{x} \) and \( \bar{y} \) are the mean values of variables x and y.

To remove the effect of magnitude in the variation of x and y, the Pearson correlation coefficient \( r \) is probably more meaningful than the covariance since it standardizes the coefficients x and y by their standard deviations:

\[
r = \frac{\text{cov}(xy)}{s_x s_y}
\]

where \( s_x \) and \( s_y \) are the standard deviations of x and y.
Fig. 3.13 Illustration of various plots with different correlation strengths (from Wonnacutt and Wonnacutt 1985 by permission of John Wiley and Sons)

**Rough thumb-rule for engg applications**

- \(\text{abs}(r)>0.9\) → strong linear correlation
- \(0.7<\text{abs}(r)<0.9\) → moderate
- \(0.7>\text{abs}(r)\) → weak

Example 3.4.2. Extension of a spring under different loads:

<table>
<thead>
<tr>
<th>Load (Newtons)</th>
<th>Extension (mm)</th>
<th>x-xbar</th>
<th>y-ybar</th>
<th>Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>10.4</td>
<td>-5</td>
<td>-24.57</td>
<td>122.85</td>
</tr>
<tr>
<td>4</td>
<td>19.6</td>
<td>-3</td>
<td>-15.37</td>
<td>46.11</td>
</tr>
<tr>
<td>6</td>
<td>29.9</td>
<td>-1</td>
<td>-5.67</td>
<td>5.07</td>
</tr>
<tr>
<td>8</td>
<td>42.2</td>
<td>1</td>
<td>7.23</td>
<td>7.23</td>
</tr>
<tr>
<td>10</td>
<td>49.2</td>
<td>3</td>
<td>14.23</td>
<td>42.69</td>
</tr>
<tr>
<td>12</td>
<td>58.5</td>
<td>5</td>
<td>23.53</td>
<td>117.65</td>
</tr>
</tbody>
</table>

Mean 7.000 34.967

**Standard deviations of load and extension are 3.742 and 18.298 respectively, while the correlation coefficient = 0.998. This indicates a very strong positive correlation between the two variables as one should expect.**

\[
\text{cov}(xy) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})
\]

<table>
<thead>
<tr>
<th>Load (Newtons)</th>
<th>Extension (mm)</th>
<th>x-xbar</th>
<th>y-ybar</th>
<th>Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>10.4</td>
<td>-5</td>
<td>-24.57</td>
<td>122.85</td>
</tr>
<tr>
<td>4</td>
<td>19.6</td>
<td>-3</td>
<td>-15.37</td>
<td>46.11</td>
</tr>
<tr>
<td>6</td>
<td>29.9</td>
<td>-1</td>
<td>-5.67</td>
<td>5.07</td>
</tr>
<tr>
<td>8</td>
<td>42.2</td>
<td>1</td>
<td>7.23</td>
<td>7.23</td>
</tr>
<tr>
<td>10</td>
<td>49.2</td>
<td>3</td>
<td>14.23</td>
<td>42.69</td>
</tr>
<tr>
<td>12</td>
<td>58.5</td>
<td>5</td>
<td>23.53</td>
<td>117.65</td>
</tr>
</tbody>
</table>

Mean 7.000 34.967

**Standard deviations of load and extension are 3.742 and 18.298 respectively, while the correlation coefficient = 0.998. This indicates a very strong positive correlation between the two variables as one should expect.**

<table>
<thead>
<tr>
<th>Load (Newtons)</th>
<th>Extension (mm)</th>
<th>x-xbar</th>
<th>y-ybar</th>
<th>Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>10.4</td>
<td>-5</td>
<td>-24.57</td>
<td>122.85</td>
</tr>
<tr>
<td>4</td>
<td>19.6</td>
<td>-3</td>
<td>-15.37</td>
<td>46.11</td>
</tr>
<tr>
<td>6</td>
<td>29.9</td>
<td>-1</td>
<td>-5.67</td>
<td>5.07</td>
</tr>
<tr>
<td>8</td>
<td>42.2</td>
<td>1</td>
<td>7.23</td>
<td>7.23</td>
</tr>
<tr>
<td>10</td>
<td>49.2</td>
<td>3</td>
<td>14.23</td>
<td>42.69</td>
</tr>
<tr>
<td>12</td>
<td>58.5</td>
<td>5</td>
<td>23.53</td>
<td>117.65</td>
</tr>
</tbody>
</table>

Mean 7.000 34.967

**Standard deviations of load and extension are 3.742 and 18.298 respectively, while the correlation coefficient = 0.998. This indicates a very strong positive correlation between the two variables as one should expect.**
3.4g Descriptive Measures of Sample Data

3.4.3 Data transformations

(a) Decimal scaling moves the decimal point but still preserves most of the original data. The specific observations of a given variable may be divided by where is the minimum value so that all the observations are scaled between -1 and 1. For example, say the largest value is 289 and the smallest value is -150, then since x=3, all observations are divided by 1000 so as to lie between [0.289 and -0.150].

(b) Min-max scaling allows for better distribution of observations over the range of variation than does decimal scaling. It does this by redistributing the values to lie between [-1 and 1]. Hence, each observation is normalized as follows:

\[ z_i = \frac{x_i - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \]

Note that though this transformation may look very appealing, the scaling relies largely on the minimum and maximum values, which are generally not very robust and may be error prone.

(c) Standard deviation scaling is widely used for distance measures (such as in multivariate statistical analysis) but transforms data into a form unrecognizable from the original data. Here, each observation is normalized as follows:

\[ z_i = \frac{x_i - \bar{x}}{s_x} \]

3.5a Plotting Data

Graphical representations of data are the backbone of exploratory data analysis. They can serve as mediums to communicate information (what the author wishes to convey) not just to explore data trends. They are usually limited to one-, two- and three-dimensional data.

In the last few decades, there has been a dramatic increase in the types of graphical displays largely due to the seminal contributions of Tukey (1988), Cleveland (1985) and Tufte (1990, 2001).

A particular graph is selected based on its ability to emphasize certain characteristics or behavior of 1-D data, or to indicate relations between 2- and 3-dimension data.

A simple manner of separating these characteristics is to view them as being:

- cross-sectional (i.e., the sequence in which the data has been collected is not retained),
- time series data,
- hybrid or combined, and
- relational (i.e., emphasizing the joint variation of two or more variables).
3.5b Plotting Data

Table 3.5.1 Type and function of graph message determines format (downloaded from http://www.eia.doe.gov/neic/graphs/introduc.htm)

<table>
<thead>
<tr>
<th>Type of Message</th>
<th>Function</th>
<th>Typical Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>Component</td>
<td>Shows relative size of various parts of a whole</td>
<td>--Pie chart (for 1 or 2 important components) --Bar chart --Dot chart --Line chart</td>
</tr>
<tr>
<td>Relative Amounts</td>
<td>Ranks items according to size, impact, degree, etc.</td>
<td>--Bar chart --Line chart --Dot chart</td>
</tr>
<tr>
<td>Time Series</td>
<td>Shows variation over time</td>
<td>--Bar chart (for few intervals) --Line chart</td>
</tr>
<tr>
<td>Frequency</td>
<td>Shows frequency of distribution among certain intervals</td>
<td>--Histogram --Line chart --Box-and-Whisker</td>
</tr>
<tr>
<td>Correlation</td>
<td>Shows how changes in one set of data is related to another set of data</td>
<td>--Paired bar --Line chart --Scatter diagram</td>
</tr>
</tbody>
</table>

3.5c Plotting Data-
univariate

Fig. 3.14 Box and whisker plot and its association with a normal distribution. The box represents the 50th percentile range while the whiskers extend 1.5 times the inter-quartile range (IQR) on either side (from Wickepedia website).
3.5d Plotting Data- previous univariate example

Suggestions on number of bins for histograms:

(a) Devore and Fornum (2005) suggest: 
\[ N_{\text{bins}} = (n)^{1/2} \]
which would suggest that if \( n = 100 \), \( N_{\text{bins}} = 10 \)

(b) Doebelin (1995)
\[ N_{\text{bins}} = 1.87.(n-1)^{0.4} \]
which would suggest that if \( n = 100 \), \( N_{\text{bins}} = 12 \).
3.5f Plotting Data- bivariate and multivariate

Fig. 3.17 Two different ways of plotting stationary data. Data corresponds to worldwide percentages of total primary energy supply in 2003 (from IEA, World Energy Outlook, IEA, Paris, France, 2004).

3.5g Plotting Data- bivariate and multivariate

Fig. 3.18 Different types of bar plots to illustrate year by year variation (over 6 years) in quarterly electricity sales (in GigaWatt-hours) for a certain city.
3.5h Plotting Data

Fig. 3.19 Scatter plot (or x-y plot) with trend line through the observations. In this case, a second order quadratic regression model has been selected as the trend line.

3.5i Plotting Data

(a) Low resolution

(b) High resolution

Fig. 3.20 Figure to illustrate how the effect of resolution can mislead visually. The same data is plotted in the two plots but one would erroneously conclude that there is more data scatter around the trend line for (b) than for (a).
3.5j Plotting Data- *Dot plots*

Fig. 3.21 Commute patterns in major U.S. cities in 2008 shown as enhanced dot plots with the size of the dot representing the number of commuters (from Wikipedia website)

3.5k Plotting Data- *Combination Plots*

Fig. 3.22 Several types of combination charts are possible. The plots shown allow visual comparison of the standardized (subtracted by the mean and divided by the standard deviation) hourly whole-house electricity use in a large number of residences against the standard normal distribution (from Reddy 1990)
3.5 l Plotting Data- **Combination Plots**

Fig. 3.23 Scatter plot combined with box-whisker-mean (BWM) plot of the same data as shown in Fig. 3.11 (from Haberl and Abbas 1998 by permission of Haberl)

Fig. 3.24 Example of a combined box-whisker-component plot depicting how hourly energy use varies with hour of day during a year for different outdoor temperature bins for a large commercial building

3.5m Plotting Data- 3-D Plots

Fig. 3.25 Three dimensional surface charts of mean hourly whole-house electricity during different hours of the day across a large number of residences (from Reddy 1990)

Fig. 3.26 Example of a three-dimensional plots of measured hourly electricity use in a commercial building over nine months
3.5n Plotting Data- **Contour Plots**

![Contour Plot]

Fig. 3.27 Contour plot characterizing the sensitivity of total power consumption (condenser water pump power plus cooling tower fan power) to condenser water-loop controls for a single chiller load, ambient wet-bulb temperature and chilled water supply temperature.

3.5o Plotting Data- **Carpet Plots**

![Scatter Plot Matrix]

Fig. 3.29 Scatter plot matrix or carpet plots for multivariable graphical data analysis. The data corresponds to hourly climatic data for Phoenix, AZ for January 1990. The bottom left hand corner frame indicates how solar radiation in Btu/hr-ft\(^2\) (x-axis) varies with dry-bulb temperature (in °F) and is a flipped and rotated image of that at the top right hand corner. The HR variable represents humidity ratio (in lbm/lb). Points which fall distinctively outside the general scatter can be flagged as outliers.
3.5p Plotting Data- Graphical Treatment of Outliers

Fig. 3.30 Illustration of different types of outliers. Point A is very probably a doubtful point; point B might be bad but could potentially be a very important point in terms of revealing unexpected behavior; point C is close enough to the general trend and should be retained until more data is collected.

Fig. 3.31 Two other examples of outlier points. While the outlier point in (a) is most probably a valid point, it is not clear for the outlier point in (b). Either more data has to be collected, failing which it is advisable to delete this data from any subsequent analysis (from Belsley et al. 1980 by permission of John Wiley and Sons).

3.6 Overall Measurement Uncertainty

The International Organization of Standardization (ISO) and six other organizations have published guides which have established the experimental uncertainty standard (an example is ANSI/ASME, 1990).

3.6.1 Need for uncertainty analysis

Any measurement exhibits some difference between the measured value and the true value and, therefore, has an associated uncertainty. A statement of measured value without an accompanying uncertainty statement has limited meaning.

*Uncertainty* is the interval around the measured value within which the true value is expected to fall with some stated confidence level.

“Good data” does not describe data that yields the desired answer. It describes data that yields a result within an acceptable uncertainty interval or, in other words, provides the acceptable degree of confidence in the result.
3.6 Overall Measurement Uncertainty

Measurements made in the field are especially subject to potential errors. In contrast to measurements made under the controlled conditions of a laboratory setting, field measurements are:

- typically made under less predictable circumstances
- with less accurate and less expensive instrumentation
- errors in variable measurement conditions so that the method employed may not be the best choice for all conditions
  - errors due to limited instrument field calibration, because it is typically more complex and expensive than laboratory calibration
  - errors due to simplified data sampling and archiving methods
  - limitations in the ability to adjust instruments in the field.

Fig. 3.32 Effect of measurement bias and precision errors

- **Random error**: due to the unpredictable nature of errors, can be treated by statistical methods, error of this type reduces with the number of readings
- **Bias error**: systematic error (due to instrument or the way it is placed)-statistics is of limited use, error of this type does not reduce as more readings are taken
3.6.3 Random Uncertainty

- UNCERTAINTY: interval around the measured value within which the true value is expected to fall at some confidence interval
- For example, the statement that the measurement is within 5.1 - 8.2 at 95% implies that the probability of the actual value being between {5.1, 8.2} is 95%
- A higher confidence interval would result in a wider range and vice versa
- Usual confidence levels (CL) used are: 99%, 95% and 90%.
- Significance level is also used:
  - for two-tailed distribution, 95% CL is the same as 0.025 significance level

Random errors can be:
- **Additive**: independent of magnitude of reading
  - say, instrument has an error of 5% of full scale
- **Multiplicative**: dependent on magnitude of reading
  - say, instrument has an error of 5% of measured value
Multiple Measurements Assuming Random

To determine uncertainty bands for “n” measurements of the same quantity- we use probability tables

Most commonly used:
- Normal or Gaussian distribution table
- Student t-distribution table for smaller samples

- For Z=1, one would expect 68.3% of data will be within (1 x std) of mean
- For Z=2, 95.5% of the data will be within (2 x std) of mean, and
- For Z=3, 99.7% of the data will be within (3 x std) of the mean

- Look at the t- table to see how the degrees of freedom widen the uncertainty bands

Random uncertainty of large samples (n > 30)

The uncertainty level is given by:

\[
\begin{align*}
\bar{x}_{\text{min}} &= \bar{x} - \left( \frac{Zs_x}{\sqrt{n}} \right) \\
\bar{x}_{\text{max}} &= \bar{x} + \left( \frac{Zs_x}{\sqrt{n}} \right)
\end{align*}
\]

Z is a multiplier deduced from the normal distribution
- For 95.0% CL, Z=1.96
- For 99.0% CL, Z=2.58

Random uncertainty of small samples (n < 30)

\[
\begin{align*}
\bar{x}_{\text{min}} &= \bar{x} - \left( \frac{t s_x}{\sqrt{n}} \right) \\
\bar{x}_{\text{max}} &= \bar{x} + \left( \frac{t s_x}{\sqrt{n}} \right)
\end{align*}
\]

t is a multiplier deduced from the Student t-tables for n-1 degrees of freedom
- For 95.0% CL and d.f.=10, t=2.228
- For 99.0% CL and d.f.=10, t=3.169
Example 3.6.1. Estimating confidence intervals

(a) The length of a field is measured 50 times. The mean is 30 with a standard deviation of 3. Determine the 95% CL.

This is a large sample case, for which the $z$ multiplier is 1.96. Hence, the 95% CL are:

$$30 \pm \frac{(1.96)(3)}{\sqrt{50}} = 30 \pm 0.83 = \{29.17, 30.83\}$$

(b) Only 21 measurements are taken and the same mean and standard deviation are found. Determine the 95% CL.

This is a small sample case for which the $t$-value=2.086 for d.f.=20. Then, the 95% CL will turn out to be wider:

$$30 \pm \frac{(2.086)(3)}{\sqrt{21}} = 30 \pm 1.37 = \{28.63, 31.37\}$$
Overall Uncertainty

- Earlier expression for uncertainty
  \[ \delta_i = \beta + \varepsilon_i \]
  where:
  \( \delta_i \) = total measurement error
  \( \beta \) = bias or fixed error
  \( \varepsilon_i \) = random or precision error
  \( i \) = a subscript to denote a particular experiment

- Current
  \[ U_x = \sqrt{B_x^2 + \left( \frac{s_i}{\sqrt{n}} \right)^2} \]

  \( U_x \) = overall uncertainty in the value \( x \) at a specified confidence level
  \( B_x \) = uncertainty in the bias or fixed component at the specified confidence level
  \( s_i \) = standard deviation estimates for the random component
  \( n \) = sample size
  \( t \) = \( t \)-value at the specified confidence level for the appropriate degrees of freedom

Example 3.6.2: For a single measurement, the statistical concept of standard deviation does not apply. Nonetheless, one could estimate it from manufacturer’s specifications if available.

To estimate the overall uncertainty at 95% confidence level in an individual measurement of water flow rate in a pipe under the following conditions:
- full scale meter reading 150 L/s
- actual flow reading 125 L/s
- random error of instrument is ±6% of full-scale reading at 95% CL
- fixed (bias) error of instrument is ±4% of full-scale reading at 95% CL

The solution is rather simple since all stated uncertainties are at 95% CL. It is implicitly assumed that the normal distribution applies.

The random error = 150x0.06 = ±9 L/s.
The fixed error = 150x0.04 = ±6 L/s
The overall uncertainty can be estimated from Eq. 3.18 with \( n=1 \):

\[ U_x = (6^2 + 9^2)^{1/2} = 10.82 \text{ L/s} \]

The fractional overall uncertainty at 95% CL = \( \frac{U_x}{x} = \frac{10.82}{125} = 0.087 = 8.7\% \)
Example 3.6.3. Consider Example 3.6.2. In an effort to reduce the overall uncertainty, 25 readings of the flow are taken instead of only one reading. The resulting uncertainty in this case is determined as follows.

- The bias error remains unchanged at ±6 L/s.
- The random error decreases by a factor of \( \sqrt{25} = 5 \)
  \[ \frac{9}{25} \] to \( \pm 1.8 \) L/s
- The overall uncertainty is thus:
  \[ U_x = (6^2 + 1.8^2)^{1/2} = \pm 6.26 \text{ L/s} \]

The fractional overall uncertainty at 95% confidence level is:

\[ \frac{U_x}{x} = \frac{6.26}{125} = 0.05 = 5.0\% \]

Increasing the number of readings from 1 to 25 reduces the relative uncertainty in the flow measurement from ±8.7% to ±5.0%. Because of the large fixed error, further increase in the number of readings would result in only a small reduction in the overall uncertainty.

Example 3.6.4 A flow meter manufacturer stipulates a random error of 5% for his meter at 95.5% CL. Once installed, the engineer estimates that the bias error due to the placement of the meter in the flow circuit is 2% at 95.5% CL. The flow meter takes a reading every minute, but only the mean value of 15 such measurements is recorded once every 15 minutes. Estimate the uncertainty at 99% CL of the mean of the recorded values.

Bias error:

Normal table: since 95.5% CL corresponds to \( Z=2 \), we deduce the error at one standard deviation = \( 2.0/2=1.0\% \)

Normal table: 99% CL corresponds to \( Z=2.58 \)

Random error

Normal table: since 95.5% CL corresponds to \( Z=2 \), we deduce the error at one standard deviation = \( 5.0/2=2.5\% \)

Student-t table, for \( \text{d.f.}=15-1=14 \) and 99% CL- critical for \( t \) value = 2.977

Hence, the uncertainty of the recorded values at 99% CL is:

\[ \left[ 2.58 \times 1.0 \right]^2 + \left[ \frac{2.977 \times 2.5}{\sqrt{15}} \right]^2 \right]^{1/2} = 3.22\% \]
3.6.6 Chauvenet’s Criterion of Rejecting Data

Assumes normal distribution and constant variance

Points can be rejected if their deviation from the mean > d max

This max deviation is given in the table and also by:

\[
d_{\text{max}} = 0.819 + 0.544 \ln(n) - 0.02346 \ln(n^2)
\]

\[
d_{\text{max}} / \sigma
\]

where \( \sigma \) is the standard deviation and \( n \) the number of data points

Rejection criterion: probability of occurrence > (1/2n)

<table>
<thead>
<tr>
<th>Number of Readings</th>
<th>Deviation ratio ( d_{\text{max}} / \sigma )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1.15</td>
</tr>
<tr>
<td>3</td>
<td>1.38</td>
</tr>
<tr>
<td>4</td>
<td>1.54</td>
</tr>
<tr>
<td>5</td>
<td>1.65</td>
</tr>
<tr>
<td>6</td>
<td>1.73</td>
</tr>
<tr>
<td>7</td>
<td>1.8</td>
</tr>
<tr>
<td>10</td>
<td>1.96</td>
</tr>
<tr>
<td>15</td>
<td>2.13</td>
</tr>
<tr>
<td>20</td>
<td>2.31</td>
</tr>
<tr>
<td>25</td>
<td>2.33</td>
</tr>
<tr>
<td>30</td>
<td>2.51</td>
</tr>
<tr>
<td>50</td>
<td>2.57</td>
</tr>
<tr>
<td>100</td>
<td>2.81</td>
</tr>
<tr>
<td>300</td>
<td>3.14</td>
</tr>
<tr>
<td>500</td>
<td>3.29</td>
</tr>
<tr>
<td>1000</td>
<td>3.48</td>
</tr>
</tbody>
</table>

3.7 Propagation of Errors

**Addition or subtraction:**

\[
y = x_1(\pm s_{x1}) \pm x_2(\pm s_{x2})
\]

\[
y = (s_{x1}^2 + s_{x2}^2)^{1/2}
\]

**Multiplication:**

\[
s_y = (x_1 x_2) \left( \frac{(s_{x1}^2)^2 + (s_{x2}^2)^2}{x_1 x_2} \right)^{1/2}
\]

**Division:**

\[
s_y = \left( \frac{x_1}{x_2} \right) \left( \frac{(s_{x1}^2)^2 + (s_{x2}^2)^2}{x_1 x_2} \right)^{1/2}
\]

Fractional standard deviation with three variables:

\[
\frac{S_y}{y} = \left( \frac{s_{x1}^2}{x_1^2} + \frac{s_{x2}^2}{x_2^2} + \frac{s_{x3}^2}{x_3^2} \right)^{1/2}
\]
Standard deviation of a function \( y = y(x_1, x_2, ..., x_n) \), with independent measured variables all given with the same confidence level, is obtained by the first order expansion of the Taylor series:

\[
s_y = \sqrt{\sum_{i=1}^{n} \left( \frac{\partial y}{\partial x_i} s_{x,i} \right)^2}
\]  
3.7.1a

where:

- \( s_y \) = function standard deviation
- \( s_{x,i} \) = standard deviation of measured quantity \( x_i \)

If two variables \( x_1 \) and \( x_2 \) are correlated, then the standard deviation of their sum is given by:

\[
s_y^2 = s_{x1}^2 + s_{x2}^2 + 2 \cdot \text{cov}(x_1, x_2)
\]

Note: covariance can be negative

Example 3.7.1- Uncertainty in overall heat transfer coefficient

Over-all heat-transfer coefficient \( U \) of HX (neglecting thermal resistance of pipe)

\[
U=\frac{1}{h_1 + 1/h_2} = \frac{h_1 h_2}{(h_1 + h_2)}
\]

If \( h_1 = 15 \) W/m\(^2\)-\(^\circ\)C with a fractional error of 5% at 95% CL and \( h_2 = 20 \) W/m\(^2\)-\(^\circ\)C with a fractional error of 3%, also at 95% CL, what will be the fractional error in random uncertainty of \( U \) at 95% CL assuming bias error to be zero?

**Answer:** Partial derivatives computed analytically using basic calculus:

\[
\left( \frac{\partial U}{\partial h_1} \right)_{h_2} = h_2 \frac{h_1}{(h_1 + h_2)^2} = \frac{k_2^2}{(h_1 + h_2)^2}
\]  
3.27a

Expression for the fractional uncertainty in the overall heat transfer coefficient \( U \) is:

\[
\frac{S_U}{U} = \left( \frac{h_1}{(h_1 + h_2)^2} \right)^2 + \left( \frac{h_2}{(h_1 + h_2)^2} \right)^2
\]  
3.7.8

Plugging values: \( U= 8.571 \), while the partial derivatives given by Eqs. 3.27:

\[
\frac{\partial U}{\partial h_1} = 0.3265 \quad \text{and} \quad \frac{\partial U}{\partial h_2} = 0.1837
\]

Finally, \( S_U = 0.2686 \), i.e, fractional error (\( S_U / U \)) = 3.1% at 95% CL
Example 3.7.2. Relative error in Reynolds number of flow in a pipe
Adapted from Schenck, 1969

Determine the probable errors of the Reynolds numbers (Re) at the low and high flow conditions given:

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Minimum Flow</th>
<th>Maximum flow</th>
<th>Random error at full flow (95% CL)</th>
<th>% errors Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Velocity m/s (V)</td>
<td>1</td>
<td>20</td>
<td>0.1</td>
<td>10</td>
<td>0.5</td>
</tr>
<tr>
<td>Pipe diameter m (d)</td>
<td>0.2</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Density kg/m³ (ρ)</td>
<td>1000</td>
<td>1000</td>
<td>1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Viscosity kg/m-s (μ)</td>
<td>1.12x10⁻³</td>
<td>1.12x10⁻³</td>
<td>0.45x10⁻³</td>
<td>0.4</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Recall that \( Re = \frac{\rho V d}{\mu} \).

From Eq. 3.24, at minimum flow condition, the relative error in Re is:

\[
\frac{\Delta Re}{Re} = \left[ \left( \frac{0.1}{1} \right)^2 + \left( \frac{1}{1000} \right)^2 + \left( \frac{0.45}{1.12} \right)^2 \right]^{1/2} = (0.1^2 + 0.001^2 + 0.404^2)^{1/2} = 0.1 \text{ or 10%}
\]

[Note that there is no error in pipe diameter value]

At maximum flow condition, the percentage error is: 0.0065 or 0.65%

The above example reveals that
(i) at low flow conditions the error is 10% which reduces to 0.65% at high flow conditions (see Figure)
(ii) at low flow conditions the other sources of error are absolutely dwarfed by the 10% error due to flow measurement uncertainty.

Thus, the only way to improve the experiment is to improve flow measurement accuracy.
Example 3.7.3 Selecting Equipment Based on Uncertainty Calculations

The chiller cooling load at the evaporator \(Q_{ch}\) is to be monitored at an accuracy of 5%

Determined by individual measurements of the chilled water volumetric flow rate and the difference between the supply and return chilled water temperatures along with water properties.

\[
Q_{ch} = \rho V c \Delta T
\]

where:

- \(\rho = \) density of water (assumed to have no error)
- \(V = \) chilled water volumetric flow rate (kept constant)
- \(c = \) specific heat of water (assumed to have no error)
- \(\Delta T = \) temperature difference between the entering and leaving chilled water at the evaporator

Fractional uncertainty:

\[
\frac{U_{COP}}{COP} = \frac{\left(U_{L} \sqrt{V} \right)^{2}}{\left(U_{F} \right)^{2}} + \frac{\left(U_{E} \right)^{2}}{V} + \frac{\left(U_{p} \right)^{2}}{E}
\]

Note that since this is a preliminary uncertainty analysis, only random (precision) errors are considered.

1) Let us assume that the maximum flow reading of the selected meter is 1500 gpm and has 4% uncertainty at 95% CL. This leads to an absolute uncertainty of \(1500 \times 0.04 = 60\) gpm. The first term \(\frac{U_{L}}{V}\) is a constant and does not depend on the chiller load since the flow through the evaporator is maintained constant. The rated chiller flow rate is 1080 gpm, Thus \(\left(\frac{U_{L}}{V}\right)^{2} = \left(\frac{60}{1080}\right)^{2} = 0.0031\) and \(\frac{U_{L}}{V} = \pm 0.056\) .

2) Assume that for the power measurement, the instrument error at 95% CL is 4.0, calculated as 1% of the instrument full scale value of 400 kW. The chiller rated capacity is 450 tons of cooling, with an assumed realistic lower bound of 0.8 kW per ton of cooling. The anticipated electric draw at full load of the chiller = 0.8 x 450 = 360 kW. The fractional uncertainty at full load is then:

\[
\left(\frac{U_{E}}{E}\right)^{2} = \left(\frac{4.0}{360}\right)^{2} = 0.00012 \quad \text{and} \quad \frac{U_{E}}{E} = \pm 0.011
\]

The fractional uncertainty in the power is about five times less that of the flow rate.
Fractional uncertainty: \[ \frac{U_{\text{COP}}}{COP} = \sqrt{\left(\frac{U_I}{V}\right)^2 + \left(\frac{U_{\Delta T}}{\Delta T}\right)^2 + \left(\frac{U_f}{F}\right)^2} \]

3) The random (precision) error at 95% CL for the type of commercial grade sensor to be used for temperature measurement is 0.2°F. Consequently, the error in the measurement of temperature difference \( \Delta T = (0.2^2 + 0.2^2)^{1/2} = 0.28°F \). From manufacturer catalogs, the temperature difference between supply and return chilled water temperatures at full load can be assumed to be 10°F. The fractional uncertainty at full load is then

\[ \left(\frac{U_{\Delta T}}{\Delta T}\right)^2 = \frac{(0.28)^2}{10^2} = 0.00078 \quad \text{and} \quad \frac{U_{\Delta T}}{\Delta T} = \pm 0.028 \]

4) Propagation of the above errors yields the fractional uncertainty at 95% CL at full chiller load of the measured COP:

\[ \left(\frac{U_{\text{COP}}}{\text{COP}}\right) = (0.0031 + 0.00012 + 0.00078)^{1/2} = 0.063 = 6.3\% \]

Computed fractional uncertainty is not satisfactory!
The flow measurement is clearly the dominant one- try to reduce that first

Example 3.7.4. Uncertainty in exponential growth models

Exponential growth models are used to model several commonly encountered phenomena, from population growth to consumption of resources. The amount of resource consumed over time \( Q(t) \) can be modeled as:

\[ Q(t) = \int_0^t P_0 e^{rt} \, dt = \frac{P_0}{r} (e^{rt} - 1) \quad 3.32a \]

where \( P_0 \) = initial consumption rate, and \( r \) = exponential rate of growth

The world coal consumption in 1986 was equal to 5.0 billion (short) tons and the estimated recoverable reserves of coal were estimated at 1000 billion tons.

If the growth rate is assumed to be 2.7% per year, how many years will it take for the total coal reserves to be depleted?

Rearranging eq. 3.32a results in

\[ t = \frac{1}{r} \left[ \ln\left(1 + \frac{Qr}{P_0}\right) \right] \quad 3.32b \]

\[ t = \frac{1}{0.027} \ln[1 + \frac{(1000)(0.027)}{5}] = 68.75 \text{ years} \]
(b) Assume that the growth rate $r$ and the recoverable reserves are subject to random uncertainty. If the uncertainties of both quantities are taken to be normal with one standard deviation values of 0.2% (absolute) and 10% (relative) respectively, determine the lower and upper estimates of the years to depletion at the 95% confidence level.

Though the partial derivatives can be derived analytically, the use of Eq. 3.26 will be illustrated so as to compute them numerically.

Let us use Eq. 3.32b with a perturbation multiplier of 1% to both the base values of $r (=0.027)$ and of $Q (=1000)$.

<table>
<thead>
<tr>
<th>Multiplier</th>
<th>$r$ (from eq.3.32b)</th>
<th>$t$ (from eq.3.32b)</th>
<th>$Q$ (from eq.3.32b)</th>
<th>$t$ (from eq.3.32b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.99</td>
<td>0.02673</td>
<td>69.12924</td>
<td>990</td>
<td>68.43795</td>
</tr>
<tr>
<td>1.00</td>
<td>0.027</td>
<td>68.75178</td>
<td>1000</td>
<td>68.75178</td>
</tr>
<tr>
<td>1.01</td>
<td>0.02727</td>
<td>68.37917</td>
<td>1010</td>
<td>69.06297</td>
</tr>
</tbody>
</table>

$$\frac{\partial t}{\partial r} = \frac{(68.37917 - 69.12924)}{(0.02727 - 0.02673)} = -1389 \quad \text{and} \quad \frac{\partial t}{\partial Q} = \frac{(69.06297 - 68.43795)}{(1010 - 990)} = 0.03125$$

Then:

$$s_t = \left[ \left( \frac{\partial t}{\partial r} s_r \right)^2 + \left( \frac{\partial t}{\partial Q} s_Q \right)^2 \right]^{1/2}$$

$$= \left( [-1389]^{-0.002} \right)^2 + \left( [0.03125]^{-0.1} [1000] \right)^2 \right]^{1/2}$$

$$= (2.778^2 + 3.125^2)^{1/2} = 4.181$$

Thus, the lower and upper limits at the 95% CL (with the $z=1.96$) is

$$68.75 \pm (1.96)(4.181) = \{60.55, 76.94\} \quad \text{years}$$
Example 3.7.6 Using the Monte Carlo method to determine uncertainty

Table 3.12 The first few and last few calculations used to determine uncertainty in variable $t$ using the Monte Carlo method (Example 3.7.6)

<table>
<thead>
<tr>
<th>Run #</th>
<th>$Q(1000,100)$</th>
<th>$r(0.027,0.002)$</th>
<th>$t$ (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1000.0000</td>
<td>0.0270</td>
<td>66.7518</td>
</tr>
<tr>
<td>2</td>
<td>1050.8152</td>
<td>0.0297</td>
<td>72.5652</td>
</tr>
<tr>
<td>3</td>
<td>1171.6544</td>
<td>0.0269</td>
<td>73.6445</td>
</tr>
<tr>
<td>4</td>
<td>1098.2454</td>
<td>0.0284</td>
<td>73.2772</td>
</tr>
<tr>
<td>5</td>
<td>1047.5003</td>
<td>0.0201</td>
<td>69.0848</td>
</tr>
<tr>
<td>6</td>
<td>1058.0283</td>
<td>0.0247</td>
<td>67.7451</td>
</tr>
<tr>
<td>7</td>
<td>946.8644</td>
<td>0.0263</td>
<td>68.5266</td>
</tr>
<tr>
<td>8</td>
<td>1075.5269</td>
<td>0.0277</td>
<td>71.8072</td>
</tr>
<tr>
<td>9</td>
<td>967.9137</td>
<td>0.0278</td>
<td>68.6323</td>
</tr>
<tr>
<td>10</td>
<td>1194.7164</td>
<td>0.0262</td>
<td>73.3758</td>
</tr>
<tr>
<td>11</td>
<td>747.9499</td>
<td>0.0246</td>
<td>57.2155</td>
</tr>
<tr>
<td>12</td>
<td>1099.7061</td>
<td>0.0269</td>
<td>71.5707</td>
</tr>
<tr>
<td>13</td>
<td>1074.3923</td>
<td>0.0254</td>
<td>69.1221</td>
</tr>
<tr>
<td>14</td>
<td>1003.2840</td>
<td>0.0265</td>
<td>66.2233</td>
</tr>
<tr>
<td>15</td>
<td>1071.4876</td>
<td>0.0274</td>
<td>71.3437</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Run #</th>
<th>$Q(1000,100)$</th>
<th>$r(0.027,0.002)$</th>
<th>$t$ (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>983</td>
<td>1004.2355</td>
<td>0.0242</td>
<td>76.1973</td>
</tr>
<tr>
<td>984</td>
<td>956.4792</td>
<td>0.0277</td>
<td>68.1372</td>
</tr>
<tr>
<td>985</td>
<td>1001.2961</td>
<td>0.0293</td>
<td>71.3294</td>
</tr>
<tr>
<td>986</td>
<td>1009.9830</td>
<td>0.0326</td>
<td>75.7849</td>
</tr>
<tr>
<td>987</td>
<td>1033.7338</td>
<td>0.0267</td>
<td>69.4687</td>
</tr>
<tr>
<td>988</td>
<td>934.5997</td>
<td>0.0279</td>
<td>67.6444</td>
</tr>
<tr>
<td>989</td>
<td>1006.7171</td>
<td>0.0282</td>
<td>71.8201</td>
</tr>
<tr>
<td>990</td>
<td>1133.9639</td>
<td>0.0278</td>
<td>73.6712</td>
</tr>
<tr>
<td>991</td>
<td>997.0123</td>
<td>0.0252</td>
<td>68.5173</td>
</tr>
<tr>
<td>992</td>
<td>896.9897</td>
<td>0.0257</td>
<td>63.8175</td>
</tr>
<tr>
<td>993</td>
<td>1006.2361</td>
<td>0.0263</td>
<td>71.9108</td>
</tr>
<tr>
<td>994</td>
<td>1033.8220</td>
<td>0.0294</td>
<td>72.8805</td>
</tr>
<tr>
<td>995</td>
<td>1078.6051</td>
<td>0.0295</td>
<td>73.9569</td>
</tr>
<tr>
<td>996</td>
<td>1137.6546</td>
<td>0.0276</td>
<td>73.4858</td>
</tr>
<tr>
<td>997</td>
<td>950.8749</td>
<td>0.0263</td>
<td>66.9070</td>
</tr>
<tr>
<td>998</td>
<td>1023.7850</td>
<td>0.0294</td>
<td>68.7492</td>
</tr>
<tr>
<td>999</td>
<td>939.2083</td>
<td>0.0248</td>
<td>64.5932</td>
</tr>
<tr>
<td>1000</td>
<td>849.0252</td>
<td>0.0247</td>
<td>61.0231</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Value</th>
<th>Mean</th>
<th>Stdev.</th>
<th>99.9% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Q$</td>
<td>1005.0</td>
<td>0.0272</td>
<td>68.91</td>
</tr>
<tr>
<td>$r$</td>
<td>101.82</td>
<td>0.0019</td>
<td>3.919</td>
</tr>
</tbody>
</table>

Normal Probability Plot

Check to see if the random numbers generated are really normal
3.8 Planning a Non-Intrusive Field Experiment

Any experiment should be well-planned involving several rational steps (ASHRAE, 2005):

(a) Identify experimental goals and acceptable accuracy
(b) Identify variables and relationships
(c) Establish measured variables and limits
(d) Preliminary instrumentation selection
(e) Document uncertainty of each measured variable
(f) Perform preliminary uncertainty analysis
(g) Final instrument selection and methods
(h) Install instrumentation
(i) Perform initial data quality verification
(j) Collect data (pay attention to range of variability and grid spacing)
(k) Accomplish data reduction and analysis
(l) Perform final uncertainty analysis
(m) Report results

Fig. 3.38 Two different experimental designs for proper identification of the parameter (k) appearing in the model for pressure drop versus velocity of a fluid flowing through a pipe. The grid spacing shown in plot (a) is the more common one based on equal increments in the regressor variable, while that in plot (b) is likely to yield more robust estimation but would require guessing the range of variation for the pressure drop.