THE UNIVERSITY OF MANITOBA

SOME ASPECTS OF MEASUREMENT SYSTEM ANALYSIS

BY

KIM FUNG KELVIN NG

A PRACTICUM
SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

DEPARTMENT OF STATISTICS

JUNE 2002

0-612-76829-5

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION PAGE

SOME ASPECTS OF MEASUREMENT SYSTEM ANALYSIS

BY

KIM FUNG KELVIN NG

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University

of Manitoba in partial fulfillment of the requirements of the degree

of

MASTER OF SCIENCE

Abstract

A Measurement System is an important weapon for improving product quality. An accurate and precise instrument can help us to detect changes that happen in any production process. Most importantly, a well-managed measurement system also helps a company to continuously improve their production processes. In order to maintain the accuracy and precision of gauges, an understanding of how to operate the measurement system and the sources of variability arising from the measurement process is required. In this paper, we will start from the general idea of a measurement system and progress to a detailed discussion of gauge variability estimates.

Acknowledgements

First of all, I would like to give thanks to my God. He gave to me this great opportunity to come to Winnipeg for study. All His guidance and protection have led me to finish my Masters program. Secondly, I have to offer special thanks to my supervisor Professor Macpherson for his help, advice and encouragement. His effort and full commitment to students like me gave me the strength to complete this practicum. Thirdly, I have to thank my parents, my sister and my brother for their love and financial support. Fourthly, I am very appreciative for the financial support provided from Department of Statistics, and also the help, advice and constant educational experience from the Professors and staff from Department of Statistics. Finally, I wish to thank Dr. Montgomery and Dr. Nelson for their permission to use their materials in my practicum.

Contents

Chapter 1

Introduction

In today's world, customer satisfaction with a company's products is a major factor in achieving a company's success. Improving the quality of the products produced can help the company maintain market share in the global marketplace. For these reasons, the requirement for achieving the reliability of process data is a major concern of the production process. Also, the quality of data from the production process depends in large part on the instrument used in the measurement process. As a consequence, the measurement process plays a major role in ensuring that the data from a production process is accurate and precise. A reliable measurement process also helps the company to continuously improve their production processes. If the data obtained from measurement process are reliable, then improvements to the production process can be based on the information contained within these data.

In daily production, reliable measurements do not occur naturally. An understanding of the sources of variability and the ability to control the factors that affect the measurement quality in the measurement process are major components of the strategy for quality improvement. Variability in the measurement process may come from instrument calibration, parts, operators, measuring procedures, instrument repeatability and reproducibility etc. Therefore, one significant goal is to determine how much of the process variability comes from the measurement process and then to reduce and control this variability. The use of statistical techniques is helpful in monitoring the measurement process; a most important aspect is to seek ways in which the variation from the measurement process can be minimized.

The first part of this paper will focus on the introduction of the measurement system. This will include the concept of accuracy and definitions, history of measurement, general principles and practical implications, calibration, operational definitions, statistical properties, general guidelines of test procedures, and measurement control systems. The second part of this paper will focus on the gauge capability study; this will include classical gauge repeatability and reproducibility study, an experimental design approach to a gauge capability study, and construction of confidence intervals. After that, we will describe other sources of measurement error, such as gauge accuracy, linearity and stability. Also, we will describe how to compare trueness and precision for two instruments. Finally, the last part of this paper will provide an example of a gauge capability study using semiconductor data.

Chapter 2

Some Measurement Concepts

2-1 Definitions

Before we have a further discussion, some of the terms about a measurement process need to be defined. The following definitions are taken from the nomenclature documents of the *International Organization for Standardization ISO 3534-1:1993 (Statistics – Vocabulary and symbols – Part 1: Probability and general statistical terms)* and *ISO 5725-1:1994 {Accuracy (trueness and precision) of measurement methods and results – Part 1: General Principles and definitions}*.

Observed value: The value of a characteristic obtained as the result of a single observation.

Test result: The value of a characteristic obtained by carrying out a specified test method.

Level of the test in a precision experiment: The general average of the test results from all laboratories for one particular material or specimen tested.

Accepted reference value: A value that serves as an agreed-upon reference for comparison, and which is derived as:

a) a theoretical or established value, based on scientific principles;

b) an assigned or certified value, based on experimental work of some national or international organization;

c) a consensus or certified values, based on collaborative experimental work under the auspices of a scientific or engineering group;

d) when a), b) and c) are not available, the expectation of the (measurable) quantity, i.e. the mean of a specified population of measurements.

Accuracy: The closeness of agreement between a test result and the accepted reference value.

Trueness: The closeness of agreement between the average value obtained from a large series of test results and an accepted reference value.

Bias: The difference between the expectation of the test results and an accepted reference value.

Precision: The closeness of agreement between independent test results obtained under stipulated conditions.

Repeatability: Precision under repeatability conditions.

Repeatability conditions: Conditions where independent test results are obtained with the same method on identical test items in the same laboratory by the same operator using the same equipment within short intervals of time.

Repeatability limit: The value less than or equal to which the absolute difference between two test results obtained under repeatability conditions may be expected to be with a probability of 95%.

Reproducibility: Precision under reproducibility conditions.

Reproducibility conditions: Conditions where test results are obtained with the same method on identical test items in different laboratories with different operators using different equipment.

Reproducibility limit: The value less than or equal to which the absolute difference between two test results obtained under reproducibility conditions may be expected to be with a probability of 95%.

Outlier: A member of a set of values which is inconsistent with the other members of that set.

We provide two additional definitions taken from *Measurement Systems Analysis Reference Manual* (1990).

Gauge: Any device used to obtain measurements; frequently used to refer specifically to the devices used on the shop floor; includes go/no-go devices.

Measurement system: The collection of operations, procedures, gauges and other equipment, software, and personnel used to assign a number to the characteristic being measured; the complete process used to obtain measurements.

2-2 The Concept of Accuracy

In this section, we shall describe the concept of *accuracy, precision, trueness, repeatability* and *reproducibility*. For a better understanding of those concepts, we shall use some examples to explain the meaning of each. Some of the examples below are taken from Kane (1989).

The concept of accuracy can be considered to consist of two important components: trueness and precision. Understanding of these components is crucial to what follows as much of what is done in this paper is related to estimating or determining the magnitude of these factors.

Trueness, as previously defined, describes how close the long-term average of measurements made can be expected to be to the true value, or to an accepted reference value. We will obtain a measure of trueness by determining the difference between the expectation of the measurement results and the true or accepted reference value, the bias.

The second important component of accuracy, precision, refers to the closeness of agreement among measurement results. Precision then is closely linked to the variation observed between measurement observations.

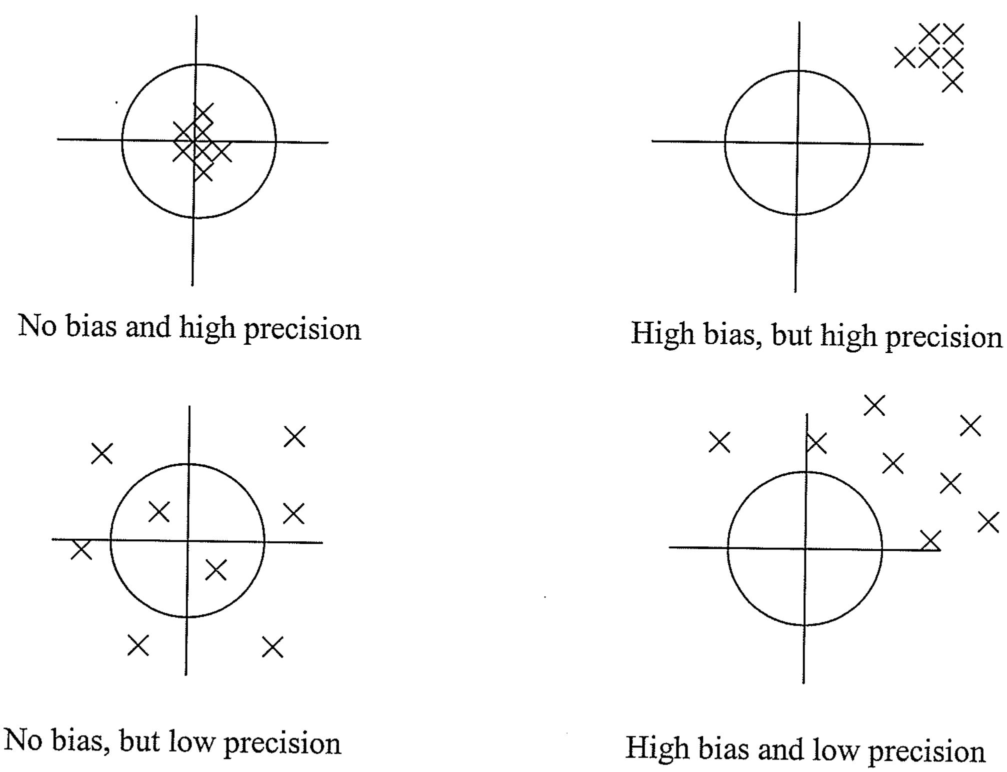


Figure 2.1 The concept of accuracy

We illustrate these ideas of accuracy by using the target diagrams given in Figure 2.1. From the target chart in Figure 2.1, we can see that the result displays high trueness if the test results are distributed around the accepted reference value (centre). Also, the result is precise if the test results are clustered closely together.

Example 2.1 (Trueness)

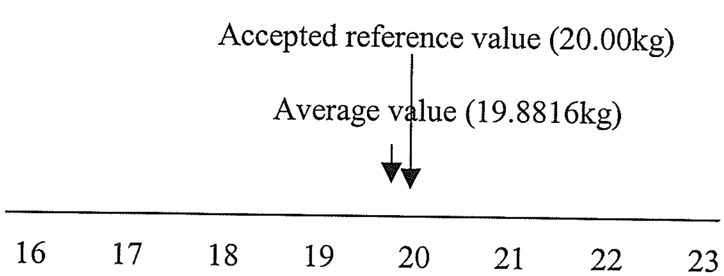
In Section 2-1, we defined trueness as the closeness of agreement between the average value obtained from a large series of test results and an accepted reference value. For example, a single part is measured repeatedly using single instrument under repeatability conditions.

The twenty-five repeated measurements on a 20-kg accepted reference standard are as follows:

19.93	**19.79**	**19.89**	**19.73**	**20.04**
20.05	**19.91**	**19.99**	**19.86**	**19.70**
19.95	**19.99**	**19.87**	**19.76**	**19.75**
19.74	**20.06**	**19.84**	**20.06**	**19.78**
19.76	**19.84**	**19.98**	**19.84**	**19.93**

Table 2.1

The average value for these twenty-five repeated measurements value is 19.8816. Based on the definition of *trueness* in Section 2-1, the concept of trueness can be displayed as follows:



The trueness in this example is the closeness of agreement between the average value (19.8816) and an accepted reference value (20.00). The bias can be estimated as the difference between the average and the accepted reference value.

There are many factors that will have an impact on the accuracy of a measuring device, method, or system. Examples of such factors include

(a) the instrument, or equipment being used;

(b) the person or persons using the equipment;

(c) the methodology used in making the measurements;

(d) the environment in which the measurement is made;

(e) the calibration and maintenance of the equipment;

(f) the time interval over which measurements are made;

(g) the material being measured.

These factors, and perhaps others, will contribute to the amount of variability that is seen in measurements made in a measurement system. We will want to try to quantify the contribution to the variability that is made by factors such as these.

There are two additional components associated with the concept of precision that are of significant important in what follows in this paper. These two concepts are repeatability and reproducibility. In studying the measurement process and the factors that influence the measurements made, particularly with respect to its precision certain conditions are associated with these two components. If the factors identified in the previous paragraph, factors (a) through (g) can be held constant, we will be operating under repeatability conditions. In other words, repeatability is the variability associated with a single operator, using a single well-maintained measuring instrument, with a standard method in a controlled environment over a short period of time on the same material.

On the other hand, reproducibility conditions are those, which occur when many of the factors previously described are allowed to vary or change. Operating under reproducibility conditions will introduce a large amount of variability into the process.

Thus reproducibility is associated with maximum variability while repeatability represents the minimum variability that might be encountered.

It should be noted that because precision is variability in general, we typically would measure it by using the statistical measure the standard deviation. We will therefore want to be able to measure the level of repeatability and reproducibility through determining the standard deviation associated with measurements made under the appropriate conditions.

To illustrate these ideas we present the following examples.

Example 2.2 (Repeatability)

In Section 2-1, we defined repeatability as precision under repeatability conditions. A single part is measured repeatedly using 2 measurement instruments under repeatability conditions with the following results. Five repeated measurements of a single part used instrument 1 are as follows:

10.0 7.1 6.5 11.1 9.2

Instrument 1:

```
      x   x          x    x     x
    _________________________________
    6    7    8    9    10   11   12
```

Five repeated measurements of the same single part used instrument 2 are also found as follows:

7.9 8.7 9.3 8.5 9.6

Instrument 2:

```
               x  xx  xx
    _________________________________
    6    7    8    9    10   11   12
```

The spread of the measurements for instrument 1 is greater than for instrument 2. Therefore, instrument 2 has better repeatability than instrument 1.

Example 2.3 (Reproducibility)

In Section 2-1, we defined reproducibility as precision under reproducibility conditions. A single part is measured repeatedly using 2 measurement instruments with 2 operators under reproducibility conditions with the following results.

	Operator 1	**Operator 2**
Instrument 1	**2.3, 1.5, 6.1, 4.2, 2.5**	**2.7, 1.9, 5.9, 6.5, 3.2**
Instrument 2	**2.7, 4.1, 3.4, 2.2, 3.1**	**6.1, 2.5, 5.2, 4.7, 5.9**

Table 2.2

Instrument 1:

```
      x  o  xxo  o     x        o x o
    _________________________________
    1    2    3    4    5    6    7
```

Instrument 2:

```
          x o x  x x     x    o    o      o o
   ____________________________________________
   1       2        3        4        5        6        7
```

The measurements from operator 1 and operator 2 using instrument 1 show greater intermixing than for instrument 2. Therefore, instrument 1 is said to be more reproducible than instrument 2.

Chapter 3

Measurement and Measurement System

3-1 A Brief History of Measurement

In the earliest stages of the world's commercial history, it didn't really matter how long was a mile or a yard. The most important fact was that each and every user of a particular measurement unit should have the same understanding of its meaning. Later, people began to develop some measurement units based on the parts of the human body. For example, the Egyptian standard unit of length *Royal Egyptian Cubit* was established about 2900 BC. It was defined as the length of the arm from the elbow to the outstretched fingertips. After this measurement unit was established, there have followed many standard units of length determined in a similar way, such as using length of arms and feet for distance measurement.

With the expansion of trading between nations, the need for seeking a universal standard of measurement was recognized. At that time, metrologists tried to develop the concept of international units of measurement. However, when the Roman Empire fell at about 600 AD, Europe then entered into the *Dark Ages*. The progress towards measurement standardization came to a virtual standstill. Then, in the thirteenth century, King Edward I of England ordered an iron stick to be made as a standard yardstick for his entire kingdom. This yardstick was called the *iron ulna*. The length of this yardstick was standardized as a yard.

In 1793, the French government adopted the *metric system* (a system of measurement standards), and the unit of length in the metric system was based on the meter. The meter was defined as one ten-millionth of a quadrant of the earth's meridian. When more accurate instruments became available, the exactness of this absolute standard became questionable. Later, some absolute standards based on an observable physical phenomenon were found, such as wavelength. Finally, this developed into the *SI system* (international system of measurement units).

The French government tried to convince the United States to convert to the metric system in 1795, but the Congress of the United States didn't take any action. In 1821, John Quincy Adams wrote a report to Congress detailing how important it was to bring uniformity in weights and measures. In 1866, Congress permitted the use of the metric system of measurement in the United States. In 1875, the United States together with seventeen other countries held an International Conference on Weights and Measures. In the same year, they signed the "Treaty of the Meter" which led to the establishment of the International Bureau of Weights and Measures based in France. Finally in 1975, Congress in the United States passed the Metric Conversion Act for adopting the *SI system* as the predominant system of measurement units to be used.

In Canada, an Act of Parliament has legalized the use of the metric system since 1871. However, the metric system was not widely used, because the use of metric system was purely voluntary. Finally, the government announced in a "White Paper" its commitment to metric conversion in 1970. In 1971, the government created *Metric*

Commission Canada for managing the conversion. In the last 20 years, the *SI System* has been taught in all Canadian schools, colleges and universities and it is now widely used in Canada. Unfortunately full conversion to the metric system has not been accomplished and *Metric Commission Canada* was disbanded in 1985.

3-2 Introduction – Measurement System

3-2.1 Why do have a measurement system?

When starting to think about the measurement system, there is a very important question for us to consider. Why do we have a need a measurement system? In other word, we have to know why the measurement system plays such a significance role in any production process. There are some reasons for us to consider.

For any company, improving the quality of products is the trademark of the company's success. In order to continue improving quality, we have to ensure that the data obtained from the entire production process is accurate and precise. To fulfill this purpose, a reliable measurement system is required. A reliable measurement system can help us to detect and identify the sources of variability in the manufacturing process and to resolve associated problems as soon as we can. Also, we can use the data that we obtained from the measurement process to determine those areas that we should strive to improve. One of the sources of variability comes from the measurement process itself. If the variation from the measurement process can be minimized, then we increase the chances that the variation from the production process can also be minimized.

3-2.2 Types of data

As we start to consider how to make measurements, we have to consider the types of data we will encounter. Data can be considered to be of the following two types:

(1) Attribute data: The data are based on occurrences of a particular feature associated with an individual part. For example, the attribute might be dents on a body panel used in the production of automobiles. The attribute data results then from the counts of the number of such panels with the given feature, in this case, dents.

Example 3.1 [This example is from Farnum (1994)]

Two auditors are each examining 100 accounting records for possible errors. The attribute in this example is the presence or absence of errors on a record. The data are the results of the determinations made for each record by the two auditors. The results of their examination are summarized in Figure 3.1.

		Auditor 2	
		T	**T***
Auditor 1	**T**	72	12
	T*	2	14
			N=100

Figure 3.1 100 accounting records for possible errors by each of two auditors

Here **T** denotes records that the auditor determines contain no errors and hence are classified as conforming. Then, **T*** denotes records that the auditor determines contain errors and hence are classified as nonconforming. The entries are the counts of the results of their examinations.

For obtaining attribute data, so-called go/no-go gauges are widely used in industry. The idea of a go/no-go gauge is basically to design a device for checking that any individual part meets a specification or not. If the part meets the specification it is classified as conforming, otherwise, the part is considered to be nonconforming. For example, a go/no-go gauge that we use on a part must be constructed to have physically the same specification as the part being tested. Then we pass the part through the ready-to-use go/no-go gauge and can determine visually whether the part is satisfactory.

Based on what we described above, we might have some ideas why go/no-go gauges are widely used in industry. This is because, the only thing operators have to do is to determine if the parts pass through the go/no-go gauge or not. Such gauges are easy to use and they don't usually require any special techniques. With a very simple and often inexpensive device the operator can determine quickly if the manufactured part has met the required specification.

In many situations however just meeting a required specification may not be sufficient. If an assembly requires the combination of several parts each part just passing specification requirements may result in an assembly that doesn't work well. The go/no-go gauge does not provide information about the variability in the process on a part-to-part basis. Nor does the classification of parts as simply conforming or non-conforming provide information about the closeness of the part to the desired target value.

(2) Measured data: the data are based on the measurement of some physical feature (e.g. furnace temperature, hole size). Measured data are recorded using some measurement scale, such as meters, degrees, or kilograms.

For obtaining measured data, we have to use a measuring instrument that may be simply a ruler or a weight scale, for example. However, in some situations for making special measurements we may have to use sophisticated instruments, such as a coordinate measuring machine (CMM).

Measured data are more informative than attribute data because the measured data allow us to describe the distribution of measurement values. For example, for a car door, it would be useful to measure and record the length and height of a car door to gain an understanding of how the manufacturing process for the doors is performing. This is more meaningful than to just count how many doors do not fit on a vehicle. However, attribute data are still widely encountered in industry because of the widespread use of go/no-go gauging.

3-2.3 Measurement system – Practical implications

In last section, we described the types of data that we would encounter. Before we start to make measurements, some practical implications of the measurement system have to be considered. The material that we will describe in what follows is taken from *ISO 5725-1:1994 {Accuracy (trueness and precision) of measurement methods and results – Part 1: General Principles and definitions}*.

(1) Standard measurement method

In every measurement process, we would like the measurement that an operator makes to be consistent. In other words, all measurements that each and every operator makes are to be done in the same way. Therefore, the measurement method has to be standardized. In order to achieve this purpose, we have to

(i) carry out the measurement according to the prescribed standard measurement method;

(ii) prepare a written document which provides in full detail how the measurement is to be made;

(iii) include a description for how to obtain and prepare the part to be measured.

(2) Accuracy experiment

For the purpose of determining the accuracy of a measurement system, we will want to construct an experiment consisting of a series of tests or measurements. The organization shall assign a panel of experts established specifically for that purpose to organize it. The estimate of accuracy is valid only when the test results are made according to the prescribed standard measurement method.

(3) Identical test items

The measurement materials or parts that are to be used in an accuracy experiment are sent from a central point. In order to have a consistent measurement, test items must remain identical during the time when the material or part is being measured. For this reason, the following two conditions have to be satisfied:

(i) The samples have to be identical when dispatched for the test.

(ii) The samples have to remain identical during the testing.

(4) Short intervals of time

According to the definition of repeatability conditions, the measurement has to be made under constant operating conditions; i.e., the measurement obtained by the same operator, under the same environment condition etc. For this reason, the time intervals for the tests under repeatability conditions have to be as short as possible. This is because we would like to minimize the changes from different factors.

(5) Observation conditions

The observed values that we obtain may be affected by the variability contributed from many factors. This variability can result from the operator, the equipment used, the calibration of the equipment, the environment, and the time elapsed between measurements. Under repeatability conditions, these factors have to be held constant when the observations are carried out. Under reproducibility condition, at least some of the factors may change when the observations are carried out. Therefore, it is very important to establish appropriate observation conditions (i.e., which factor should keep constant or which factor should be changed).

3-2.4 Measurement system – Calibration

Based on what we have described, we know how important it is for us to follow those practical implications when making measurements. However, this is not enough for us to ensure the reliability of measurement data. Being certain about the quality of measuring instrument is also of crucial importance. In order to assure the accuracy of a measuring instrument, we need first to understand the calibration process for such an instrument.

3-2.4.1 Measurement standards

The definition of calibration is provided as follows:

Calibration*:* The procedure by which a measurement standard is transferred from a higher measurement authority (i.e., a more accurate reference) to a lower one (Farnum 1994).

The national measurement standard body that is used in the United States is the National Institute of Standards and Technology (NIST). The functions of NIST include construction and maintenance of primary standards for a company's measurement instruments. For shops and test laboratories, it is not feasible for NIST to calibrate and certify the accuracy of the enormous volume of test equipment in use directly. Instead, a systematic method is used to transfer measurement standards from level to level (i.e., from a higher measurement standard to a lower one). A process of measurement standards transference is called a *hierarchy of standards* (Figure 3.2).

For a company, a primary standard is used to assure the accuracy for each measured quantity. The primary standard is used by a few highly skilled metrologists only and is often kept in an environmentally controlled room. Also, a primary standard is the company's direct link to NIST. A secondary standard is then used to calibrate the working standard, which is an intermediate reference standard between the primary standard and the working standard. A working standard is employed solely to calibrate instruments that are used in shops and laboratories. Many workmen, inspectors, and technicians might use these instruments. These people will likely have different levels of training and skill. Therefore, the design of the instruments must be tested and confirmed for different features (i.e., ruggedness, stability and fool proofing) to minimize errors from their use by human beings.

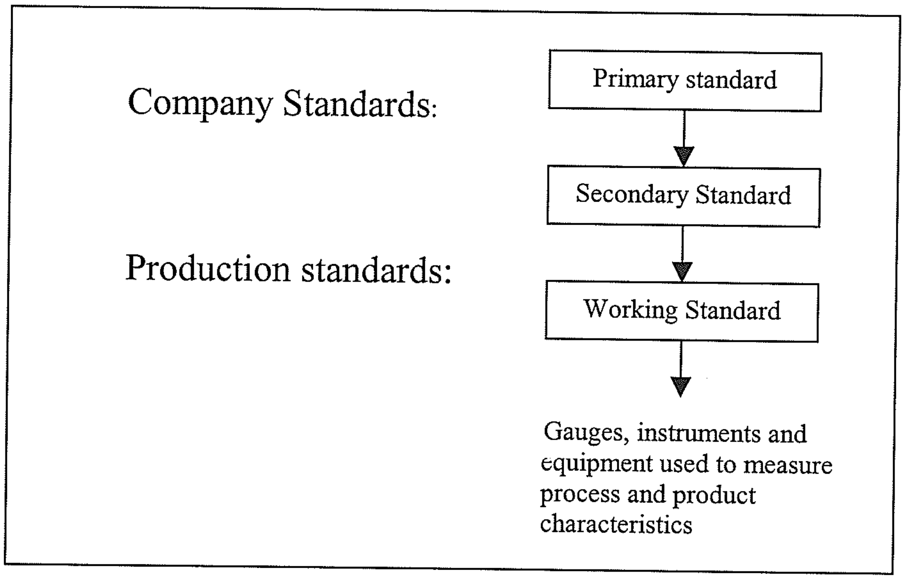


Figure 3.2 Hierarchy of standards

3-2.4.2 Calibration comparisons

Usually, we would like to compare our instrument's calibration to a higher reference standard. Two different types of calibration comparisons are as follows:

(a) Direct comparisons: we compare a measurement of an object to a measurement of a standard gauge block. For example, we compare a measurement of an object's weight to a 20-kg check standard.

(b) Indirect comparisons: this comparison involves intermediate mechanisms and calculations for converting an instrument reading (i.e., physical principles). For example, a multimeter is an instrument that is designed to use for either as an ammeter, a voltmeter, or an ohmmeter, a device to measure resistance. Physical laws are then used to calculate the resistance of a conductor (i.e., Ohm's law $V = IR$) for comparison of the measurement.

3-2.4.3 Comments on building a calibration system

Sobralske (1989) discusses the importance of a calibration system that is based on the document MIL-STD-45662A. MIL-STD-45662A is a U.S. military document that describes the requirement for creating and maintaining a calibration system used by any prime or subcontractor of the government. In his paper, he points out that the keys to compliance to MIL-STD-45662A are documentation and formalization. For a calibration system this document requires the following:

(i) All measuring and test equipment should be listed (i.e., micrometers, thermometers, gauges, voltmeters, standard, or hardness testers).

(ii) Each item on the list must be given an identification code (i.e., 10 voltmeters can be labels from V1 to V10). This identification is important because through it we can trace back information about the last calibration.

(iii) Require a documented description of a calibration system. A report should include instrument calibration procedures, type of measuring and test equipment, instrument's identification code, calibration frequency, standard used for calibrating instrument, NIST trace number, standard measurement, actual measurement reading, corrected measurement reading, date of calibration, and the calibrator's signature.

(iv) The frequency of calibration depends upon how frequently the gauge, measuring instrument, or test equipment is used. Therefore, interval of calibration can be lengthened or shortened.

(v) The environment for calibrating and using an instrument must be controlled (i.e., temperature, humidity, vibration and cleanliness).

(vi) Tags or other signals are used to warn the potential user about the status of an instrument (i.e., an instrument out of service, out of calibration, or awaiting calibration).

(vii) Devise a recall system to remind calibrator which items need checking on any given day, week, or month.

3-2.5 Measurement system – Operational definitions

We already have the idea of how to assure the accuracy of a measuring instrument. Relying on the calibration process is still not sufficient for us to achieve reliable measurement data. In order to achieve reliable measurement data, good communications between vendor, purchaser and production worker are also important. In what follows, we will discuss how all parties involved in the measurement system can be led to having the same understanding of the measurement process.

Deming (1982) commented on the meaning of the term "*operational definition*":

"An operational definition is one that people can do business with. In other words, an operational definition of safe, round, reliable, or of any other quality must be communicable, with the same meaning to vendor as to purchaser, same meaning yesterday and today to the production worker."

To conduct any type of measurement, it is important to understand the procedures of the measurement process. Misunderstanding of the measurement process by production workers may cause measurement data inconsistency (e.g. buyer and seller may be using different measurement scales). Therefore, careful specification of the measurement process is important in order to ensure the collection of data with high quality. Also, an operational definition is not simply about concepts, but it is about everything involving usage in practice (e.g. explains what measurement to make, what criteria to adopt, or listing out possible errors). For those reasons, the development of proper operational definitions is an essential part of any production or measurement process.

The development of operational definitions is very important for quality improvement. First, the resulting data should then be reliable, because operational definitions provide detailed instructions to different operators. Second, the various sources of variation in the process can be identified and addressed, and the variation coming from different operators can be minimized.

For every production process, a common way to specify the measurement procedures is provide actual examples (e.g. use of photographs to show the steps for measuring process, technical reports, or examples showing possible systematic errors). In addition, an operational definition might have to be redefined from time to time, such as to accommodate environmental condition changes.

3-2.6 Measurement system – Statistical properties

In the last section, we described how important it is for all parties involved in the measurement system to have the same understandings of the whole measurement process. In this section, we will discuss in detail how to achieve this purpose.

In 1990, three American automotive companies (Chrysler, Ford and General Motors) in association with the American Society for Quality Control developed a document called *Measurement Systems Analysis Reference Manual*. Some basic principles about the measurement system are provided in this manual. The material that we will discuss in this section is taken from this document.

The desired aim for any measurement system is to obtain "correct" measurements (e.g. the measurements agree with a target value) every time we make a measurement. However, this ideal situation is seldom if ever achieved in any measurement system. In reality, we will use statistical characteristics or properties of the measurements to describe how the measurement system is behaving. For example, if all measurements produced by a measurement system were to agree with a target value, then we could say that this measurement system has the statistical property of zero variance.

Statistical properties are used to characterize the quality of a measurement system. For each measurement system, achieving appropriate statistical properties is required. The responsibility of the facility management is then to ensure that statistical property requirements for a measurement system are met.

Some statistical properties may not be appropriate for all measurement systems. For example, suppose we desire to have a measured value that is "close" in some sense to a target value. However, the various statistical properties that could be used to characterize "close" can be quite different. We might interpret "close" to be small bias, small variance, or small type I and type II error. We must therefore specify clearly the statistical properties that are desired for each measurement system.

The properties of a system that are often of considerable interest and importance may include:

(1) The measurement system has to be in statistical control, which means that the measurement error is due solely to common causes, and not due to special causes.

(2) Variability of the measurement system must be small compared to the variability of the manufacturing production process.

(3) Variability of the measurement system must be small compared to the specification limits.

(4) The increments of the measurement device that is used for measurement system has to be no greater than one - tenth of the smaller of either the process variability or the specification limits.

(5) If different items are being measured by the measurement system, then the appropriateness of statistical properties has to be checked.

3-2.7 Measurement system – General guidelines of test procedures

Previously, we described the importance of assuring the accuracy of a measuring instrument and good communications between vendor, purchaser and production worker. However, we have to ensure that the measurement system is adequate over a period of time. In order to determine if a measurement system is adequate for the intended task, conducting one or more tests is necessary. Usually, this determination is assessed in two phases.

The phase one assessment is to conduct a test for the measurement system based on the statistical properties of a system. If a measurement system is acceptable by the test, then this system can be used in the facility. Otherwise, this measurement system has to be improved or replaced.

The phase two assessment is to retest the quality of the measurement system periodically. The reason for a retest is to verify the measurement system remains acceptable.

Documentation of the test procedures for each measurement system should include the following materials:

(1) practical examples;

(2) detail about how to select the items for the measurement process and the environment for the test procedure;

(3) detail about to how the data are to be collected, recorded, and analyzed;

(4) operational definitions required for the measurement system;

(5) if special standards (e.g. an object obtained from NIST) are to be used in the procedure, then instructions for storage, maintenance, and use of the standard have to be included.

3-2.8 Measurement system – Measurement control systems

Since measured data provides greater information about the characteristics of the production process, we will focus our attention solely on how such measured data are obtained. In order to achieve reliable measurement data, we cannot rely solely upon the quality of the measuring instrument itself. All the various procedures that are involved in the use and care of the instrument are equally important. All these factors lead us to define and discuss the entire system that is put into place to support the obtaining of measurements.

In Section 2-1, we defined measurement system as the collection of operations, procedures, gauges and other equipment, software, and personnel used to assign a number to the characteristic being measured; the complete process used to obtain measurements. In this section, we will describe more detail about one such system.

The *International Organization for Standardization (ISO)* is working on the development of an international standard describing requirements for a measurement control system. A document is being considered as a Draft International Standard, *ISO/DIS 10012 (Measurement control systems)* that will ultimately become an *ISO* International Standard. The document contains definitions and details, which will be outlined in what follows.

1) Definitions

The draft standard presents definitions for specific terms used in the document. The definitions given in the document are as follows:

Measurement control system: Set of interrelated or interacting elements necessary to achieve metrological confirmation and continual control of measurement processes.

Measurement process: Set of operations to determine the value of a quantity.

Measuring equipment: Measuring instrument, software, measurement standard, reference material or auxiliary apparatus or combination thereof necessary to realize a measurement process.

Metrological characteristic (of a measuring equipment): Distinguishing feature which can influence the results of measurement.

Metrological confirmation: Set of operations required to ensure that measuring equipment conforms to the requirements for its intended use.

Metrological function: Function with organizational responsibility for defining and implementing the measurement control system.

The *ISO/DIS 10012* document provides details of a measurement control system that could well be referenced on a contractual basis between, for example a customer and supplier, or by a government agency. In what follows, we will describe the elements of the document in order to provide the basic idea of a measurement system.

2) Management responsibility

In order to ensure that the metrological functions of the organization are prescribed and maintained, the senior management must commit sufficient resources throughout the organization.

(i) Measurement control system

The effectiveness of the measurement control system depends on how well the metrological function is managed. The management of the metrological function must establish, document, maintain and improve the effectiveness of the measurement control system continuously. Any change to the system must be according to the organization's requirements.

(ii) Customer requirements

Company's success always depends on how well the company fulfills customer requirements. The management of the metrological function is to ensure that the measurement needs and expectations of the customer are determined and converted into metrological requirements. Compliance to these requirements must be able to be demonstrated.

(iii) Quality objectives

Quality objectives for the measurement control system must be determined by the organization to indicate what it would like to achieve in its metrological system. The management of the metrological function must ensure that objective performance standards and procedures for the measurement control system have been defined and put into place. Some examples of such quality objectives are:

(1) prior to use all measuring equipment must have received confirmation;

(2) all confirmations of measuring equipment must be completed on time;

(3) no measurement process can be allowed to operate in an out-of-control state without detection for more than one day.

(iv) Management review

Management is to ensure that a systematic review of the measurement control system is conducted at regular planned intervals. There must be the resources available to assess the measurement control system. The reviews must be documented, problems identified in the review corrected, and all actions taken recorded.

3) Resource management

- Human resources

(i) Assignment of personnel

The responsibilities of all personnel assigned to the operation and control of the measurement control system must be defined and documented by the management of the metrological function.

(ii) Competence and training

The management of the metrological function has to ensure that all personnel involved in the measurement control system have enough skill and knowledge to complete their assigned tasks. If any special skills are required, it is necessary to provide a document to detail those requirements. The management must ensure that effective training is provided to address identified needs. Also, the evaluation and documentation of training is necessary, because it gives us an idea of how effective the training has been.

- Information resources

(i) Procedures

Instructions for every procedure in the measurement control system must be documented. Validation of all procedures is to be achieved to guarantee consistency of application, and validity of measurement results. Any changes to documented procedures have to be authorized and controlled.

(ii) Software

Software used in measurement process must be carefully assessed to ensure that it is suitable for continued use. All calculations of results must be documented, identified, and controlled. If there are any revisions to software, the revised software must be tested before it is placed back into continuous use.

(iii) Records

All information required for the operation of the measurement control system must be identified, documented, and maintained.

(iv) Identification

In the measurement control system, all process elements are to be clearly identified. The status of equipment with respect to confirmation must be identified, and such identification clearly attached to all equipment.

- Material resources

(i) Measuring equipment

For a measurement control system, all measuring equipment needed to assess metrological requirements must be evaluated. The results of evaluation must be identified in the measurement control system records. To avoid the invalidation of measurement results, measuring equipment must be confirmed and used in an appropriately controlled environment.

(ii) Storage and handling of measuring equipment

In order to prevent abuse, misuse, damage and changes in metrological characteristics of measuring equipment, the management of the metrological function will provide documented procedures for storage and handling of measuring equipment. Any procedures for receiving, handling, transporting, storing and dispatching measuring equipment shall be included in the documentation.

(iii) Environmental conditions

Environmental conditions in which equipment operates can have a major influence on the results obtained. The proper operating conditions must be determined, documented and monitored.

(iv) Outside suppliers

If a measurement control system requires products and services from outside suppliers, the management of the metrological function must define and document those requirements. Outside suppliers must be selected according to their ability to meet the documented requirements. Any criteria for selection, monitoring, and evaluation of suppliers must be defined and documented.

4) Measurement control system realization

- Metrological confirmation process

(i) Metrological confirmation

It is necessary to confirm that each measuring instrument or equipment is able to meet its metrological requirements. A program of metrological confirmation must be designed and implemented.

(ii) Confirmation intervals

All equipment must be confirmed at regular pre-determined intervals in order to be certain of compliance with metrological requirements on a continuous basis.

(iii) Equipment characteristics

Information about confirmation status of all equipment has to be readily available to user. Any limitations or special requirements must be documented.

(iv) Equipment adjustment control

Once a measuring device has been confirmed access to any adjustment controls on the device that can be used to change its performance must be controlled to prevent tampering. The use of seals or other such items to prevent and identify unauthorized access is important.

(v) Confirmation process records

Records must be maintained for all confirmed metrological equipment. The record must specify details of how the equipment was confirmed, and show that the equipment is able to satisfy requirements. Any and all special conditions regarding the use of the equipment must be provided in the record.

- Measurement process

(i) Process design and planning

The determination of the metrological requirements shall be based on customer, organization, and statutory and regulatory requirements. In order to meet these specified requirements, the measurement processes developed to address these requirements have to be documented.

To show that product requirements have been met, measurement processes have to be determined, planned, validated, implemented, and controlled.

(ii) Process realization

The performance characteristics required for the various measurement processes may differ depending upon their intended use. Such required performance characteristics of the measurement process must be identified and quantified. Some examples of characteristics are:

(1) repeatability;

(2) reproducibility;

(3) stability;

(4) maximum permissible error.

(iii) Nonconforming equipment

From time to time confirmed measuring equipment may have been subject to conditions which may invalidate its confirmation. Examples of such conditions affecting the equipment are

(1) damaged;

(2) overloaded;

(3) a malfunction that impacts its intended use;

(4) known or incorrect measurement results;

(5) in use beyond its confirmation interval;

(6) mishandled;

(7) access seal is damaged.

(iv) Records of control of measurement processes

In order to be able to demonstrate compliance with the elements of the measurement control system, the organization must maintain detailed records. Included in the record shall be, for example

(1) full details of the measurement process control system as implemented;

(2) data from the control system particularly with respect to measurement uncertainty;

(3) actions taken as a result of the collected data.

- Measurement realization

(i) Measurement uncertainty

Measurement uncertainty has to be estimated for each measurement process in the system. Confirmation of the measuring equipment and the validation of the measurement process shall not be approved until the analysis of measurement uncertainty is completed. Sources of variability in the measurement process are to be documented.

(ii) Traceability

The management of metrological function must ensure that all measurements are traceable in some acceptable manner to known SI units of measurement or acceptable reference standard or method. Such reference standard or method must be specified in detail and agreed to by all parties.

5) Measurement control system analysis and improvement

(i) Measurement control system audit

In order to be able to provide verification that its measurement control system is functioning in compliance with requirements, the organization will conduct planned audits of the system. Individuals not involved with the activity being audited shall conduct the audit. Deficiencies identified by the audit shall be rectified without undue delay. Results of audit and all resulting actions must be recorded.

(ii) Monitoring of measurement processes

The organization must regularly monitor the controlled measurement process according to documented procedures in order to prevent deviation from requirements. This will ensure prompt detection of deficiencies and early corrective action.

(a) Analysis of the measurement processes

For each measurement process, the particular element or characteristic to be analyzed must be determined. Limits for these characteristics must be established that reflect the risk associated with failure to comply with requirements. The analysis shall address the issue of measurement uncertainty in relation to metrological requirements.

(b) Corrective action for the measurement process

Having established an appropriate characteristic and its limits for a measurement process, corrective action must be taken when such characteristics falls outside the accepted limits or show unacceptable patterns of behavior.

(iii) Improvement

The continual improvement of the measurement control system shall be planned and managed by management.

The material that has been provided in this section described the requirements of a measurement control system. It has been included to demonstrate the importance placed on a properly maintained, documented, and functioning system. In what follows in the paper we will consider many of the statistical and analytical tools that would be useful in such a measurement control system.

Chapter 4

Gauge Capability Studies

Montgomery and Runger (1993a) stated:

"The goal of a gauge capability analysis is the understanding and quantification of the sources of variability present in the measurement process."

For these reasons, a successful gauge capability study based upon proper experimental principles is a very important component for improving the quality of the measurement process. Also, a well-designed gauge capability study will

(1) provide good and reliable estimates of the variation in the measurement process;

(2) identify those factors that are most influential with respect to the inherent variation of the process.

In what follows, we will describe some of the aspects of planning gauge capability studies. Then, we will discuss different types of gauge capability studies.

4-1 Planning Gauge Capability Studies

4-1.1 Selection of parts

For a gauge capability study, the number of parts used in the measurement process is an important consideration for measurement analysis. Some people prefer to make repeated measurement on few parts. However, Montgomery and Runger (1993a) have argued that there are some advantages to using many parts in a gauge capability study.

(1) A collection of parts selected from the production process provides a broad coverage of the materials on which the measurement system actually operates. Measurement of a "standard" unit near the centre of the manufacturing specifications might show less variation than measurement made at the extremes of manufacturing specifications. In order to increase the chance of detecting this situation, we should use many parts that span the entire specification range.

(2) Non-constant measurement variance might not be detected if only a few parts are involved in the study (i.e., measurement proportional to the mean level of the product). Therefore, non-constant measurement variance can potentially be detected when using parts that span a broad measurement range. Visual inspection of the data can help to detect non-constant variance. An R chart, constructed from measurement data ordered according to the mean of the measurements on a part, may reveal patterns suggesting a lack of consistent variability.

(3) When multiple measurements are made, the operator is often required to make repeated measurements on the same part. Usually, the part is mounted with some holding device, the instrument zeroed, and the measurement made. In making repeated measurements on the part, the danger is that the operator may simply re-measure the part repeatedly without removing it after each measurement. In this situation, when repeated measurements are to be made, the operator is less likely to perform a complete replication of the measurement process. A complete replication

of the measurement process requires that the part be mounted, the instrument zeroed, the measurement made, and the part then is removed after the measurement is made. This process of mounting, zeroing, measuring and removing is then repeated for each measurement on the same part. Every part is measured in this exact same way. If the operator does not perform complete replication of the measurement process, then some of the important sources of variation might not be taken into account. This is because different sources of variation have greater opportunity to occur when measuring a new part than when making repeated measurements on the same part. Using more parts therefore may be better because it forces the operator to do a complete replication of the process.

For these reasons, parts selection should try to span as much of the range of potential measurements as is possible. However, it is possible for a manufacturing process to produce unusual parts at low frequency. It is then likely that a sample of 20 parts, say, will not include any of these unusual parts. Therefore, we should supplement our sample of parts with historical parts from the extremes of the manufacturing process. If historical parts are not available, we would supplement our sample with standards at the extremes of manufacturing process.

4-1.2 Accelerating long-term variability

For the gauge capability study, some of the sources of variability might only appear after a long period of time. These sources of variability give rise to what is called *Long-term variability.* Because the cost of a measurement capability study must be controlled, it might not be possible economically to include these sources of variability directly. Therefore, we would like to devise some methods to enable the estimation of the long-term sources of variability in the measurement process. The usual way is to identify the factors that maybe considered to be possible sources of long-term variability and to attempt to include them when conducting the gauge capability study (e.g. include different gauges used in different locations, different environmental conditions, or different operators, etc.). Montgomery and Runger (1993a) discussed a simple approach for conducting the gauge capability study for the situation that we mentioned above.

This approach is to conduct a preliminary analysis for the beginning of a gauge capability study. This preliminary analysis would try to include as many factors, that might contribute to the variation for the measurement process, as is possible. For example, we can measure the same part using different operators and equipment in different locations. Also, using different combinations of operators, equipment and location on another part might simulate different sources of variability. This preliminary study, being very simple and unstructured, provides an easy way to obtain preliminary information for a gauge capability study. If the result of such a preliminary study is acceptable, then the measurement process can also be considered to be acceptable. This is because, a preliminary analysis of this kind can be considered to represent a worst-case evaluation of gauge capability analysis.

However, when the results show that the measurement process is not capable of meeting the requirement of the manufacturing process, this approach is not satisfactory. Such an unstructured preliminary study lacks the structure of a well-designed experiment and hence there is no way to tell which factors and interactions are important. As a result, it is very difficult, if not impossible, to determine where the unacceptable measurement variability comes from.

4-2 A Simple Gauge Capability Study

A gauge capability study is conducted in order for us to determine how well, or how badly the measurement gauge is performing. A simple capability study will involve an operator taking repeated measurements on a collection of parts selected from the production process. From the variability observed in the measurements made, it will be possible to identify and to estimate the variability that is attributed to the parts, and to the measurement instrument. We will thus be able to partition the observed total variation into its components. In a simple experiment, for example, the total variability can be expressed as follows:

$$\sigma^2_{total} = \sigma^2_{product} + \sigma^2_{gauge} \tag{4.1}$$

where σ^2_{total} is the total variance, $\sigma^2_{product}$ is the variance component for the product, and σ^2_{gauge} is the variance component for the gauge. Using the data from the experiment, we will be able to obtain estimates of these variances. As part of a gauge capability analysis, it is helpful to use graphical data displays, such as X-bar and R charts. To illustrate we describe a simple gauge capability study. In this example, a single operator measured 25 different parts using a single instrument and each part was measured twice. In order to construct the X-bar and R charts, we calculate the average and the range for the two measurements of each part. The data are taken from Farnum (1994) and displayed in Table 4.1.

Example 4.1

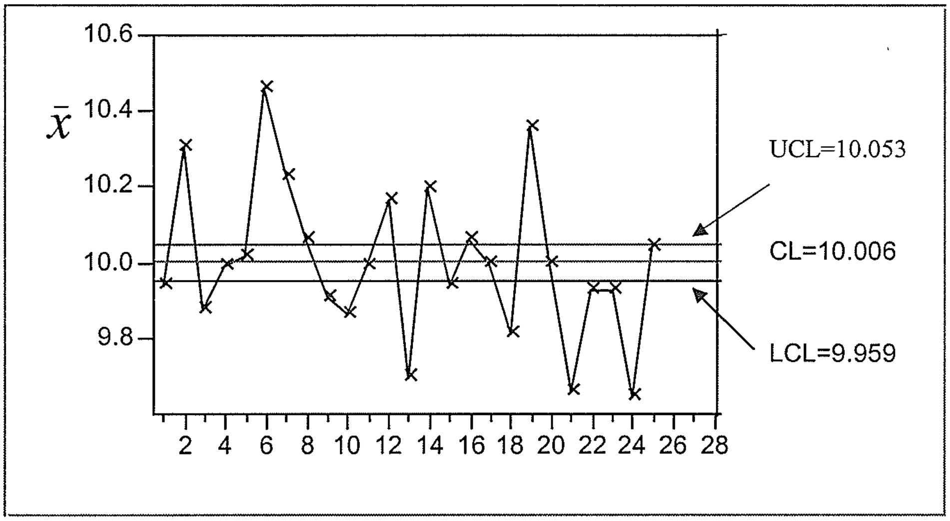
The data are displayed as follows:

	Measurements			
Part	1	2	$\bar{x}$	R
1	9.966	9.925	9.946	0.041
2	10.319	10.298	10.309	0.021
3	9.874	9.878	9.876	0.004
4	10.014	9.975	9.995	0.039
5	10.009	10.027	10.018	0.018
6	10.488	10.430	10.459	0.058
7	10.249	10.208	10.229	0.041
8	10.064	10.060	10.062	0.004
9	9.913	9.908	9.911	0.005
10	9.836	9.901	9.869	0.065
11	10.018	9.967	9.993	0.051
12	10.165	10.172	10.169	0.007
13	9.707	9.696	9.702	0.011
14	10.211	10.184	10.198	0.027
15	9.931	9.959	9.945	0.028
16	10.081	10.052	10.067	0.029
17	9.981	10.014	9.998	0.033
18	9.787	9.847	9.817	0.060
19	10.352	10.372	10.362	0.020
20	9.995	10.004	10.000	0.009
21	9.655	9.661	9.658	0.006
22	9.926	9.937	9.932	0.011
23	9.925	9.934	9.930	0.009
24	9.653	9.638	9.646	0.015
25	10.040	10.053	10.047	0.013
			$\bar{\bar{x}} = 10.006$	$\bar{R} = 0.025$

Table 4.1

The X-bar and R charts for these data are as follows:

X-bar Chart:



R chart:

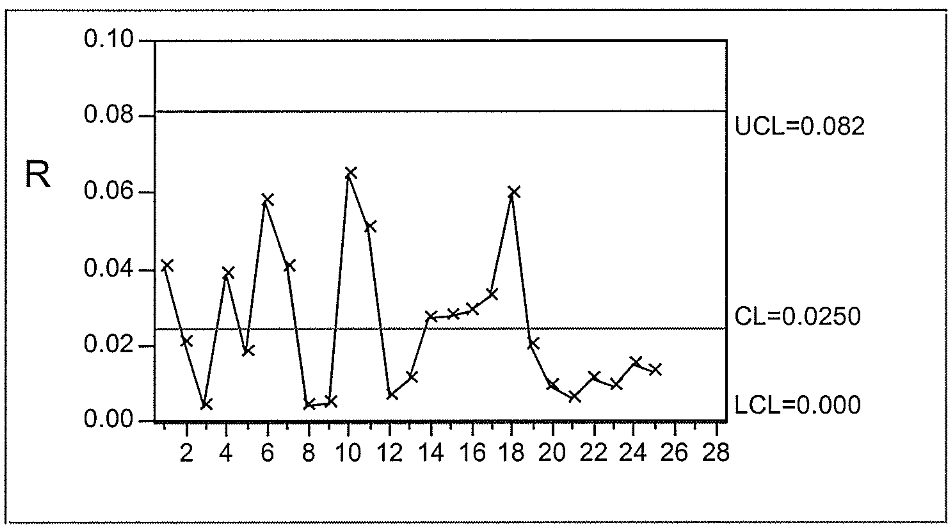


Figure 4.1

The R chart is constructed to allow us to examine the variability within the repeated measurements made on a single part. The R chart in Figure 4.1 shows no points plotting outside of the control limits. This indicates that the operator is making consistent measurements. On the other hand, X-bar chart shows many out of control points. On the surface this may seem surprising. It is crucial however to recall how the control limits for the X-bar chart are determined and what the chart is actually displaying. In this example, the X-bar chart is a plot of the average of the two measurements made on each part. Thus the chart is showing the observed part-to-part variability. The calculation of control limits on the X-bar chart is based upon the average of the Ranges of the repeated measurements on a single part however. As a result, the Ranges reflect the observed within part variability due to measurement error. It would be anticipated that the magnitude of the within part variability will be small relative to the between part variability arising from its production process. Thus the distance between the control limits on the X-bar chart will be narrow. For this reason, we would expect that many of the averages would plot outside the control limits. This is showing that this measuring instrument is able to detect the fact that different parts have been used.

The data obtained from the sample study (Table 4.1) can be used to estimate the variance components that we described. The standard deviation associated with the gauge, σ_{gauge} can be estimated from the observed Ranges of the repeated measurements made on each of the parts:

$$\hat{\sigma}_{gauge} = \frac{\overline{R}}{d_2} = \frac{0.025}{1.128} = 0.022$$

The d_2 value is obtained from Table I in the Appendix with $n = 2$ reflecting the fact that two repeated measurements were obtained.

Because the R chart shows consistency of measurements we are able to estimate the total variance σ^2_{total} using the observed variance for all measurements.

The estimate of the variance of total variability is

$$\hat{\sigma}^2_{total} = S^2 = \frac{\sum_{i=1}^{p}\sum_{j=1}^{n}(x_{ij} - \bar{\bar{x}})^2}{(n \times p) - 1}$$

$$= \frac{(9.966 - 10.006)^2 + \cdots\cdots + (10.053 - 10.006)^2}{50 - 1}$$

$$= \frac{1.989}{49} = 0.041$$

Because the total variance is the sum of the gauge variance and the product variance, using Equation (4.1), from these results we can estimate the variance associated with the product.

The estimate of the product variance is then

$$\hat{\sigma}^2_{product} = \hat{\sigma}^2_{total} - \hat{\sigma}^2_{gauge}$$

$$= 0.041 - (0.022)^2 = 0.0405$$

So the estimated product standard deviation is

$$\hat{\sigma}_{product} = \sqrt{0.0405} = 0.201$$

4-3 The Classical Gauge Repeatability and Reproducibility Study

In Section 4-2, we discussed how to use observed ranges to estimate the standard deviations of the gauge, total and product. In this section, we will have more discussion on gauge variability. In general, gauge variability σ^2_{gauge} can be decomposed into two components of measurement error; one is the *repeatability* of the gauge and another is the *reproducibility* of the gauge. The definitions of repeatability and reproducibility and a discussion of these concepts are given in Chapter 2. The measurement error of gauge could be expressed as follows:

$$\sigma^2_{gauge} = \sigma^2_{repeatability} + \sigma^2_{reproducibility} \tag{4.2}$$

The idea of classical gauge R&R analysis is to use sample ranges to estimate the standard deviations of repeatability and reproducibility. We now use an example to illustrate the procedures of a classical gauge repeatability and reproducibility analysis.

Example 4.2

Three operators measured 25 different parts using a single instrument and each part was measured twice. The repeated measurements made by each operator are thus made under repeatability conditions. The inclusion of different operators provides us with the reproducibility conditions. In order to estimate the standard deviations of repeatability and reproducibility, we will calculate the average $\overline{x}$ and the range R for each part. The data are taken from Farnum (1994) and measurement results are listed in Table 4.2.

In this example, the three operators are each making repeated measurements on the same part using the same instrument. We thus have repeatability conditions and the differences between the repeated measurements, the sample ranges, can be used to estimate the repeatability component.

Part	Operator 1				Operator 2				Operator 3			
	Measurements				Measurements				Measurements			
	1	2	$\bar{x}_1$	R_1	1	2	$\bar{x}_2$	R_2	1	2	$\bar{x}_3$	R_3
1	10.210	10.260	10.235	0.050	10.150	10.190	10.170	0.040	10.160	10.170	10.165	0.010
2	9.830	9.790	9.810	0.040	9.800	9.760	9.780	0.040	9.770	9.800	9.785	0.030
3	10.050	10.060	10.055	0.010	9.920	9.920	9.920	0.000	10.100	10.090	10.095	0.010
4	10.010	10.020	10.015	0.010	10.000	9.970	9.985	0.030	9.980	9.970	9.975	0.010
5	9.800	9.850	9.825	0.050	9.840	9.820	9.830	0.020	9.780	9.790	9.785	0.010
6	10.120	10.170	10.145	0.050	10.120	10.100	10.110	0.020	10.100	10.090	10.095	0.010
7	10.230	10.260	10.245	0.030	10.260	10.230	10.245	0.030	10.240	10.290	10.265	0.050
8	10.190	10.180	10.185	0.010	10.130	10.130	10.130	0.000	10.160	10.150	10.155	0.010
9	10.000	10.040	10.020	0.040	9.980	10.020	10.000	0.040	10.060	10.020	10.040	0.040
10	9.710	9.700	9.705	0.010	9.680	9.720	9.700	0.040	9.730	9.740	9.735	0.010
11	10.240	10.280	10.260	0.040	10.100	10.140	10.120	0.040	10.200	10.220	10.210	0.020
12	10.060	10.010	10.035	0.050	10.070	10.080	10.075	0.010	10.110	10.120	10.115	0.010
13	10.020	10.060	10.040	0.040	10.020	10.030	10.025	0.010	10.080	10.090	10.085	0.010
14	9.870	9.890	9.880	0.020	10.030	10.050	10.040	0.020	9.870	9.870	9.870	0.000
15	9.830	9.920	9.875	0.090	9.950	9.950	9.950	0.000	9.880	9.940	9.910	0.060
16	10.360	10.330	10.345	0.030	10.270	10.300	10.285	0.030	10.350	10.340	10.345	0.010
17	10.150	10.100	10.125	0.050	10.160	10.180	10.170	0.020	10.260	10.210	10.235	0.050
18	9.840	9.860	9.850	0.020	9.850	9.830	9.840	0.020	9.860	9.890	9.875	0.030
19	10.580	10.590	10.585	0.010	10.500	10.470	10.485	0.030	10.470	10.500	10.485	0.030
20	9.980	10.020	10.000	0.040	10.040	10.000	10.020	0.040	9.990	10.020	10.005	0.030
21	10.030	10.020	10.025	0.010	9.990	9.970	9.980	0.020	9.970	9.970	9.970	0.000
22	9.700	9.660	9.680	0.040	9.760	9.700	9.730	0.060	9.780	9.760	9.770	0.020
23	10.150	10.140	10.145	0.010	10.090	10.050	10.070	0.040	10.080	10.130	10.105	0.050
24	9.980	10.000	9.990	0.020	9.980	9.980	9.980	0.000	9.950	9.980	9.965	0.030
25	10.220	10.210	10.215	0.010	10.150	10.190	10.170	0.040	10.170	10.160	10.165	0.010
	$\bar{\bar{x}}_1 = 10.052$ $\bar{R}_1 = 0.031$				$\bar{\bar{x}}_2 = 10.032$ $\bar{R}_2 = 0.026$				$\bar{\bar{x}}_3 = 10.048$ $\bar{R}_3 = 0.022$			

Table 4.2 Data for Example 4.2

For the estimation of gauge repeatability, we will use the average of the three average ranges:

$$\overline{\overline{R}} = \frac{1}{3}(\overline{R}_1 + \overline{R}_2 + \overline{R}_3)$$

$$= \frac{1}{3}(0.031 + 0.026 + 0.022)$$

$$= 0.026$$

The average range, $\overline{\overline{R}}$, reflects the within part variability due to measurement error. The standard deviation associated with the gauge repeatability, $\sigma_{repeatability}$ can be estimated from the average range $\overline{\overline{R}}$:

$$\hat{\sigma}_{repeatability} = \frac{\overline{\overline{R}}}{d_2}$$

$$= \frac{0.026}{1.128} = 0.023$$

The d_2 value is obtained from the Table I in the Appendix with $n = 2$ reflecting the fact that two repeated measurements were obtained by each operator.

To address the reproducibility of the gauge, we note that the three different operators measure the same part. The degree to which these measurements differ will be an indication of the reproducibility capability of the gauge. We will therefore base our estimation of the gauge reproducibility on the observed difference between the average of the measurements made by the operators.

Let $\bar{\bar{x}}_i$ be the average of all measurements made by operator i, $i = 1,2,3$. Then,

a. $\bar{\bar{x}}_{\max} = \max(\bar{\bar{x}}_1, \bar{\bar{x}}_2, \bar{\bar{x}}_3)$

b. $\bar{\bar{x}}_{\min} = \min(\bar{\bar{x}}_1, \bar{\bar{x}}_2, \bar{\bar{x}}_3)$

c. $R_{\bar{\bar{x}}} = \bar{\bar{x}}_{\max} - \bar{\bar{x}}_{\min}$

The range $R_{\bar{\bar{x}}}$ reflects the between part variability due to operators. The standard deviation associated with the gauge reproducibility, $\sigma_{reproducibility}$ can be estimated from the range $R_{\bar{\bar{x}}}$:

$$\hat{\sigma}_{reproducibility} = \frac{R_{\bar{\bar{x}}}}{d_2}$$

From Table 4.2, the estimated values of the average, $\bar{\bar{x}}_i$, by the three operators are as follows:

$$\bar{\bar{x}}_1 = 10.052,\ \bar{\bar{x}}_2 = 10.032,\ \bar{\bar{x}}_3 = 10.048$$

As a result we obtain,

$$\bar{\bar{x}}_{\max} = \max(10.052,\ 10.032,\ 10.048) = 10.052$$

$$\bar{\bar{x}}_{\min} = \min(10.052,\ 10.032,\ 10.048) = 10.032$$

$$R_{\bar{\bar{x}}} = 10.052 - 10.032 = 0.020$$

and hence

$$\hat{\sigma}_{reproducibility} = \frac{0.020}{1.693} = 0.012$$

The d_2 value is obtained from Table I in the Appendix with $n = 3$ reflecting the fact that three operators are involved in this experiment, and the range of these 3 averages is $R_{\bar{\bar{x}}}$.

Now, we can use the estimated repeatability standard deviation, $\hat{\sigma}_{repeatability}$, and reproducibility standard deviation, $\hat{\sigma}_{reproducibility}$, to estimate the standard deviation associated with the gauge, $\hat{\sigma}_{gauge}$.

Recall Equation (4.2),

$$\sigma^2_{gauge} = \sigma^2_{repeatability} + \sigma^2_{reproducibility}$$

The variance associated with the gauge can be then estimated as

$$\hat{\sigma}^2_{gauge} = \hat{\sigma}^2_{repeatability} + \hat{\sigma}^2_{reproducibility}$$

$$= (0.023)^2 + (0.012)^2 = 0.000673$$

and therefore, the gauge standard deviation estimate is

$$\hat{\sigma}_{gauge} = \sqrt{0.000673} = 0.026$$

From this estimate of the gauge standard deviation, we can obtain a measure of what is referred to as the gauge capability. Gauge capability is determined as $6\sigma_{gauge}$ and reflects the length of the interval within which virtually all repeated measurements of a part could be expected to fall. In our example, the estimated gauge capability is $6\hat{\sigma}_{gauge} = 6(0.026) = 0.156$. In the next section, we shall study the ratio between overall gauge capability $(6\sigma_{gauge})$ and tolerance band (USL – LSL), called the P/T ratio.

In this example, we can conclude that most of the measurement error was contributed from repeated measurements by different operators. This indicates that some of the operators may have difficulty in making consistent measurements and perhaps an improved gauge might be sought.

4-4 Precision-to-Tolerance Ratio (P/T ratio)

In many situations the part that is measured will have associated with it an upper specification limit (USL) and a lower specification limit (LSL) within which the manufactured part dimension is to fall. The difference between the USL and the LSL, $USL - LSL$, is referred to as the tolerance band, T.

The ratio of the gauge capability, $6\sigma_{gauge}$, to the tolerance band $T = USL - LSL$ is called the Precision-to-Tolerance Ratio, or the P/T ratio. The P/T ratio is used to compare the estimate of gauge capability $(6\hat{\sigma}_{gauge})$ to the tolerance band $(USL - LSL)$ for the part. Based on this ratio, we can examine the adequacy of the gauge that is being used in the measurement process. If the value of the P/T ratio less than or equal to 0.1, then the gauge capability may usually be considered to be adequate. This is based on the generally used criterion that the measuring instrument is able to measure in units one-tenth as large as the accuracy required in the measurement of the part.

The estimated P/T ratio can be represented as follows:

$$\frac{P}{T} = \frac{6\hat{\sigma}_{gauge}}{USL - LSL} \tag{4.3}$$

Example 4.3

In Example 4.2, the parts used in measurement has $USL = 11$ and $LSL = 9$. The P/T ratio is equal to

$$\frac{P}{T} = \frac{6(0.026)}{11-9} = \frac{0.156}{2} = 0.078$$

Values of 0.1 or less are frequently considered adequate gauge capability. In this example, the value of P/T ratio is less than 0.1; therefore, we can consider we have adequate gauge capability.

One of the difficulties of the P/T ratio is that it is dependent upon the specification limits. It is particularly noted that changing the specification limits will have a major impact on the P/T ratio. Two parts with different tolerance bands using the same measuring instrument may result in markedly different P/T ratios.

As we mentioned the P/T ratio can be affected by the width of the tolerance bands. An alternative expression for the estimate of gauge capability that does not depend upon the specification limits is given by the ratio of $\hat{\sigma}_{gauge}$ to $\hat{\sigma}_{product}$, or $\hat{\sigma}_{gauge}$ to $\hat{\sigma}_{total}$. Clearly, these ratios do not depend on the width of the tolerance band. These ratios are represented as follows:

$$\frac{\hat{\sigma}_{gauge}}{\hat{\sigma}_{product}} \times 100 \qquad (4.4)$$

and

$$\frac{\hat{\sigma}_{gauge}}{\hat{\sigma}_{total}} \times 100 \qquad (4.5)$$

Example 4.4

In Example 4.1, $\hat{\sigma}_{gauge} = 0.022$, $\hat{\sigma}_{product} = 0.201$, and $\hat{\sigma}_{total} = 0.202$. Therefore,

$$\frac{\hat{\sigma}_{gauge}}{\hat{\sigma}_{product}} \times 100 = \frac{0.022}{0.201} \times 100 = 10.945\%$$

and

$$\frac{\hat{\sigma}_{gauge}}{\hat{\sigma}_{total}} \times 100 = \frac{0.022}{0.202} \times 100 = 10.891\%$$

We can conclude that the gauge variability is about 11% of the product variability and total variability.

4-5 Experimental Design Approach for the Gauge Capability Study

In previous sections, we have provided a simple approach to addressing the issue of gauge capability through gauge repeatability and reproducibility studies. The studies that have been illustrated have used very straightforward experiments to enable us to obtain estimates of variances and standard deviations for the gauge, and product.

In attempting to analyze a measurement system, there are many factors that can have an impact on the measurements that are made. It seems natural to want to attempt to assess the amount that each of the various factors contributes to the overall variability. One approach for doing so is through the use of more elaborate designed experiments that will incorporate the factors that we may wish to study. The designed experiment will enable us to better understand how various factors influence the measurement process.

One advantage of designing an experiment is that various factors can be included in the experiment. The inclusion of operators, locations, days, instruments and the like can be helpful in gaining an understanding of the measurement process. In addition, the use of statistical models associated with designed experiments is common and well documented.

A particularly useful class of experiments that is frequently employed is the class of factorial designs. Such designs employ a set of factors with each factor included at various levels, typically but not exclusively, two levels. Given the structure of the factorial experiment we can determine how many experimental runs would be required to investigate all factors at their different levels. For example, if there are two factors and each factor has two levels, then four runs would be required for a complete trial. An example of a simple two factors each at two levels factorial model would be:

$$X_{ij} = \mu + A_i + B_j + \varepsilon_{ij} \qquad i = 1,2, \;\; j = 1,2$$

For such a model various assumption concerning the factors and the error term can be made which will enable the estimation of factor effects, variance components, and the like.

To illustrate the use of the experimental design approach, we consider the following two examples from the literature.

Example 1: Model from Montgomery and Runger for a gauge capability study

Montgomery and Runger (1993a) described an experiment in which 3 operators each measure 20 parts, with each part being measured twice. They assumed that the measurement made by operator i on part j at replication k could be modeled as a two-factor factorial model. For a test result X_{ijk}, the model can be expressed as follows:

$$X_{ijk} = \mu + O_i + P_j + (OP)_{ij} + R_{(ij)k} \qquad \begin{cases} i = 1, 2, \cdots, o \\ j = 1, 2, \cdots, p \\ k = 1, 2, \cdots, n \end{cases} \tag{4.6}$$

where μ is the overall mean, O_i is a random variable representing the effects of Operators, P_j is a random variable representing the effects of Parts, $(OP)_{ij}$ is a random variable representing the effects of Operator by Part interaction, and $R_{(ij)k}$ is a random variable representing the effects of Replications on the measurement. The Replication factor represents the Repeatability aspect of the experiment and it is nested within Operator and Part. We will provide more discussion on the gauge capability study later in this section.

Example 2: Model from Borror, Montgomery and Runger for the gauge capability study

Borror, Montgomery and Runger (1997) considered an experiment that involved measurements made at different sites by different shifts (operators) on different days. For this they developed a factorial model with nested factors. They recognized in this experiment that operators might change during the day. For this reason, the factor Shift should be nested within the factor Day. Since this structure would occur at each site, the Site factor is crossed with the nesting of Shift and Day. Also, Day would be a blocking factor in this model, and they felt there would be no interaction between Day with Site or

Shift. For every test result X_{ijkm} is the m^{th} measurement made at the k^{th} site by the j^{th} shift (operator) on the i^{th} day. The statistical model can be expressed as follows:

$$X_{ijkm} = \mu + D_i + O(D)_{j(i)} + S_k + SO(D)_{kj(i)} + R_{(ijk)m} \qquad \begin{cases} i = 1,2,\cdots,d \\ j = 1,2,\cdots,o \\ k = 1,2,\cdots,s \\ m = 1,2,\cdots,n \end{cases} \tag{4.7}$$

where μ is the overall mean, D_i is a random variable representing the Day effect, $O(D)_{j(i)}$ is a random variable representing the Shift within Day effect, S_k is a random variable representing the Site effect, $SO(D)_{kj(i)}$ is a random variable representing the Site crossed with the Shift within Day effect, and $R_{(ijk)m}$ is a random variable representing the replication effect.

When an experiment is designed to study the measurement process, many factors may be included. Some of the factors may have a major impact on the measured results while others may have little or no effect. One important aspect of the analysis of the data from a complex experiment is the assessment of the significance of the various factors and their interactions.

Using the well-known analysis of variance approach, we will be able to determine if a factor is statistically important, or if the factor can be ignored. Having determined which factors are statistically significant, we will then estimate the contribution of each of these significant factors to the overall variation of the measurements.

Models such as have been described in these examples can provide great insight into the structure of the measurement process. Moreover, we are able to use many computer programs for the analysis of the data obtained to provide estimates of the effects and variance components.

It should be noted that various assumptions must be applied to such models. For example we might assume

(1) the terms in the model have mean zero which would imply that the measurement system is unbiased;

(2) the variance components are constant;

(3) the measurement made is linear with respect to the factors included in the model.

In order to have better understanding on experiment design approach, we will use the two-factor factorial model (4.6) described above from Montgomery and Runger (1993a). This two-factor factorial model can be expressed as follows:

$$X_{ijk} = \mu + O_i + P_j + (OP)_{ij} + R_{(ij)k} \qquad \begin{cases} i = 1, 2, \cdots, o \\ j = 1, 2, \cdots, p \\ k = 1, 2, \cdots, n \end{cases}$$

Each test result X_{ijk} represents the measurement made by operator i on part j at replication k.

In many gauge capability studies, we will often assume that the parts and operators were randomly selected from a larger population of parts and operators. Therefore, the parts and operators would be treated as random effects. However, the operators used may in fact be the only operators available in some industries. In this situation, we would treat the operators as a fixed effect. In this section, we will discuss the analysis of the data under both random effects model and mixed effects model for this two-factor factorial design.

To start the discussion on the statistical analysis for the models, we need to address how to calculate all of the sum of squares associated with effects in the model, such as the total sum of squares, sum of squares for the parts, sum of squares for the operators, and so on. After the sums of squares are obtained and their corresponding degrees of freedom are determined, we will use these to calculate the mean squares of the effects. The mean squares will be used to determine the test statistics for each effect. Under appropriate assumptions, the ratios of mean squares (i.e. MS_O / MS_{OP}, MS_p / MS_{OP}, and so on) are distributed as the F-distribution. In general, the test statistics of the analysis of variance are used to help us to determine which main effects or interactions are statistically significant. An analysis of variance table as in Table 4.3 is used to summarize the analysis.

The Total sum of squares is computed as follows:

$$SS_T = \sum_{i=1}^{o}\sum_{j=1}^{p}\sum_{k=1}^{n} X_{ijk}^2 - \frac{X_{...}^2}{opn} \tag{4.8}$$

where

$$X_{...} = \sum_{i=1}^{o}\sum_{j=1}^{p}\sum_{k=1}^{n} X_{ijk} \tag{4.9}$$

and $X_{...}$ denote the grand total of all observations.

The sum of squares for the Parts is given as follows:

$$SS_{Part} = \frac{1}{on}\sum_{j=1}^{p} X_{\cdot j\cdot}^2 - \frac{X_{...}^2}{opn} \tag{4.10}$$

where

$$X_{\cdot j \cdot} = \sum_{i=1}^{o} \sum_{k=1}^{n} X_{ijk} \tag{4.11}$$

and $X_{\cdot j \cdot}$ denote the total of all observations associated with the j^{th} Part.

The sum of squares for the Operators is obtained as follows:

$$SS_{Operator} = \frac{1}{pn} \sum_{i=1}^{o} X_{i \cdot \cdot}^2 - \frac{X_{\cdots}^2}{opn} \tag{4.12}$$

where

$$X_{i \cdot \cdot} = \sum_{j=1}^{p} \sum_{k=1}^{n} X_{ijk} \tag{4.13}$$

and $X_{i \cdot \cdot}$ denote the total of all observations for made by the i^{th} Operator.

The sums of squares for the Operator by Part interaction are given as follows:

$$SS_{Interaction} = \frac{1}{n} \sum_{i=1}^{o} \sum_{j=1}^{p} X_{ij \cdot}^2 - \frac{X_{\cdots}^2}{opn} - SS_{Part} - SS_{Operator} \tag{4.14}$$

where

$$X_{ij \cdot} = \sum_{k=1}^{n} X_{ijk} \tag{4.15}$$

and $X_{ij \cdot}$ denote the total of all observations in the ij^{th} cell.

The Error sum of squares is then obtained by subtraction:

$$SS_R = SS_T - SS_{Part} - SS_{Operator} - SS_{Interaction} \tag{4.16}$$

These sums of squares and the associated degrees of freedom are displayed in Table 4.3. The entries in the column "Mean Square" are obtained by dividing each sum of squares by its degrees of freedom.

We have included various factors and interactions in the model for this experiment. For example, we have included a number of operators, several parts, and the interaction between parts and operators. We now must determine if the data obtained provide sufficient evidence that the effects associated with these factors are statistically significant. To do this we need to develop appropriate test statistics.

(1) Random Effects Model

We here will treat parts and operators as random factors since they are selected at random from a larger population. The model parameters O_i, P_j, $(OP)_{ij}$ and $R_{(ij)k}$ are thus considered to be independent random variables. We also assume that these random variables are normally distributed with mean zero and their variances are given by $\mathrm{var}(O_i) = \sigma_O^2, \mathrm{var}(P_j) = \sigma_P^2, \mathrm{var}[(OP)_{ij}] = \sigma_{OP}^2$ and $\mathrm{var}(R_{(ij)k}) = \sigma_R^2$.

To determine the test statistics for the effects and interactions, we have to examine the expected mean squares for each component. Under the assumptions made for the model, the expected mean squares for the components are given by

$$E(MS_O) = \sigma_R^2 + n\sigma_{OP}^2 + pn\sigma_O^2 \tag{4.17}$$

$$E(MS_P) = \sigma_R^2 + n\sigma_{OP}^2 + on\sigma_P^2 \tag{4.18}$$

$$E(MS_{OP}) = \sigma_R^2 + n\sigma_{OP}^2 \tag{4.19}$$

and

$$E(MS_R) = \sigma_R^2 \tag{4.20}$$

Based on the expected mean squares, we can determine the statistics for hypothesis testing of the variance components. For example, suppose we are interested in testing the hypothesis $H_0 : \sigma_O^2 = 0$. We note that under the null hypothesis $H_0 : \sigma_O^2 = 0$, the expected mean square for operators has the same form as the expected mean square for the Operator by Part interaction. This can be represented as

$$E(MS_O) = \sigma_R^2 + n\sigma_{OP}^2 \tag{4.17a}$$

In this case, the ratio of the $E(MS_O)$ in Equation (4.17a) to the $E(MS_{OP})$ in Equation (4.19) will be 1. If however σ_O^2 is not zero, then this ratio will be expected to be greater than 1. This suggests that a test statistic for testing this null hypothesis would be the ratio of the mean square MS_O to MS_{OP}. Therefore, the test statistics for testing $H_0 : \sigma_O^2 = 0$ can be expressed as follows:

$$F_0 = \frac{MS_O}{MS_{OP}} \tag{4.21}$$

and under an assumption of normality, this ratio F_0 is distributed as $F_{(o-1),(o-1)(p-1)}$.

The test statistics for testing $H_0 : \sigma_P^2 = 0$ and $H_0 : \sigma_{OP}^2 = 0$ can be determined in the same way. The statistics for testing $H_0 : \sigma_P^2 = 0$ is

$$F_0 = \frac{MS_P}{MS_{OP}} \tag{4.22}$$

and the ratio F_0 is distributed as $F_{(p-1),(o-1)(p-1)}$.

The statistics for testing $H_0 : \sigma_{OP}^2 = 0$ is

$$F_0 = \frac{MS_{OP}}{MS_R} \tag{4.23}$$

where again the ratio F_0 is distributed as $F_{(o-1)(p-1),\, op(n-1)}$.

We can also estimate the variance components by using the observed mean squares. To construct estimators for the variance components, we again consider the expected mean square provided earlier. From Equations (4.17) and (4.19), we note that

$$E(MS_O) - E(MS_{OP}) = pn\sigma_O^2$$

and hence σ_O^2 can be written as

$$\sigma_O^2 = \frac{E(MS_O) - E(MS_{OP})}{pn}$$

This suggests that an estimator of σ_O^2 is given by

$$\hat{\sigma}_O^2 = \frac{MS_O - MS_{OP}}{pn} \tag{4.24}$$

Similarly, we find the estimators for the variance components σ_P^2, σ_{OP}^2 and σ_R^2 as follows:

$$\hat{\sigma}_P^2 = \frac{MS_P - MS_{OP}}{on} \tag{4.25}$$

$$\hat{\sigma}_{OP}^2 = \frac{MS_{OP} - MS_R}{n} \tag{4.26}$$

and

$$\hat{\sigma}_R^2 = MS_R \tag{4.27}$$

The analysis of variance for the gauge capability study (random effects model) is summarized as follows:

SOURCE OF VARIABILITY	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F-Ratio
Operators	$SS_{Operator}$	$o-1$	$MS_O = \dfrac{SS_{Operator}}{o-1}$	$F_0 = \dfrac{MS_O}{MS_{OP}}$
Parts	SS_{Part}	$p-1$	$MS_P = \dfrac{SS_{Part}}{p-1}$	$F_0 = \dfrac{MS_P}{MS_{OP}}$
Operator by Part	$SS_{Interaction}$	$(o-1)(p-1)$	$MS_{OP} = \dfrac{SS_{Interaction}}{(o-1)(p-1)}$	$F_0 = \dfrac{MS_{OP}}{MS_R}$
Repeatability	SS_R	$op(n-1)$	$MS_R = \dfrac{SS_R}{op(n-1)}$	
Total	SS_T	$opn-1$		

Table 4.3

We now consider the issue of the repeatability and reproducibility as provided by the gauge capability experiment described by Montgomery and Runger (1993a). As noted previously, the replication factor considers the repeated measurements made on each part by every operator. Thus the component σ_R^2 and its estimator $\hat{\sigma}_R^2$ can here be taken to be $\sigma^2_{repeatability}$ and its estimator $\hat{\sigma}^2_{repeatability}$. The variance component of the gauge repeatability can then be estimated by

$$\hat{\sigma}^2_{repeatability} = \hat{\sigma}_R^2 = MS_R \tag{4.28}$$

In considering the reproducibility of the gauge, we again will look at the structure of the experiment. Clearly since each operator measures every part, the variability associated with operators is a component of reproducibility. However, because there is a possibility that some operators may have difficulty measuring particular parts but not others, there could be a substantial interaction between parts and operators. This interaction then contributes to the measurement error and should be considered as a component of the reproducibility. It is reasonable then to define the variance component of gauge reproducibility as

$$\sigma^2_{reproducibility} = \sigma^2_O + \sigma^2_{OP} \tag{4.29}$$

And this variance component can be estimated as

$$\hat{\sigma}^2_{reproducibility} = \hat{\sigma}^2_O + \hat{\sigma}^2_{OP} \tag{4.30}$$

Recall Equations (4.24) and (4.26),

$$\hat{\sigma}^2_O = \frac{MS_O - MS_{OP}}{pn}, \quad \hat{\sigma}^2_{OP} = \frac{MS_{OP} - MS_R}{n}$$

and hence, we obtain

$$\begin{aligned}\hat{\sigma}^2_{reproducibility} &= \frac{MS_O - MS_{OP}}{pn} + \frac{MS_{OP} - MS_R}{n} \\ &= \frac{MS_O + MS_{OP}(p-1) - pMS_R}{pn}\end{aligned}$$

Example 4.5

We now consider the gauge measurement data in Table 4.2. In Example 4.2, we showed how to use ranges $\overline{\overline{R}}$ and $R_{\overline{\overline{x}}}$ to obtain the estimation of gauge repeatability and reproducibility. In example, 3 operators measured 25 different parts using a single instrument and each part was measured twice. We can assume that this experiment could be modeled as a two-factor factorial model. The model can be represented in the same form as we previously described,

$$X_{ijk} = \mu + O_i + P_j + (OP)_{ij} + R_{(ij)k} \qquad \begin{cases} i = 1,2,3 \\ j = 1,2,\cdots,25 \\ k = 1,2 \end{cases} \tag{4.31}$$

In this example, we will treat parts and operators as random factor. We now apply the analysis of variance method to analyze the model (4.31) using the data from Table 4.2. The sums of squares are computed as follows:

$$\begin{aligned} SS_T &= \sum_{i=1}^{o}\sum_{j=1}^{p}\sum_{k=1}^{n} X_{ijk}^2 - \frac{X_{...}^2}{opn} \\ &= 15137.850 - \frac{2269873.692}{150} \\ &= 5.359 \end{aligned}$$

$$\begin{aligned} SS_{Part} &= \frac{1}{on}\sum_{j=1}^{p} X_{\cdot j \cdot}^2 - \frac{X_{...}^2}{opn} \\ &= \frac{90825.794}{6} - \frac{2269873.692}{150} \\ &= 5.141 \end{aligned}$$

$$\begin{aligned} SS_{Operator} &= \frac{1}{pn}\sum_{i=1}^{o} X_{i\cdot\cdot}^2 - \frac{X_{...}^2}{opn} \\ &= \frac{756625.089}{50} - \frac{2269873.692}{150} \\ &= 0.011 \end{aligned}$$

$$SS_{Interaction} = \frac{1}{n}\sum_{i=1}^{o}\sum_{j=1}^{p} X_{ij\cdot}^{2} - \frac{X_{\cdots}^{2}}{opn} - SS_{Part} - SS_{Operator}$$
$$= \frac{30275.624}{2} - \frac{2269873.692}{150} - 5.141 - 0.011$$
$$= 0.169$$

and

$$SS_R = SS_T - SS_{Part} - SS_{Operator} - SS_{Interaction}$$
$$= 5.359 - 5.141 - 0.011 - 0.169$$
$$= 0.038$$

The analysis of variance is shown as follows:

Source	DF	Sum of Squares	Mean Square	F-value	P-value
Operators	2	0.011	0.0055	1.57	0.2185
Parts	24	5.141	0.2142	61.20	< 0.0001
Operator by Part	48	0.169	0.0035	7.00	< 0.0001
Repeatability (R)	75	0.038	0.0005		
Total	149	5.359			

Table 4.4

Since, the P-value for Operator by Part interaction is less than 0.0001, we conclude that there is a significant interaction between Operators and Parts. The main effect of Part is also significant, however, the main effect of Operator is not statistically significant.

The gauge repeatability and reproducibility are estimated as follows:

$$\hat{\sigma}_{repeatability}^{2} = \hat{\sigma}_{R}^{2} = MS_R$$
$$= 0.0005$$

and hence

$$\hat{\sigma}_{repeatability} = \sqrt{0.0005}$$
$$= 0.0224$$

and

$$\hat{\sigma}^2_{reproducibility} = \hat{\sigma}^2_O + \hat{\sigma}^2_{OP}$$
$$= \frac{MS_O - MS_{OP}}{pn} + \frac{MS_{OP} - MS_R}{n}$$
$$= \frac{0.0055 - 0.0035}{50} + \frac{0.0035 - 0.0005}{2}$$
$$= 0.0015.$$

Hence

$$\hat{\sigma}_{reproducibility} = \sqrt{0.0015}$$
$$= 0.0387$$

If we compare the estimation of gauge repeatability and reproducibility in this analysis with that in Example 4.2, we can see that the resulting gauge repeatability estimates are similar ($\hat{\sigma}_{repeatability}$ in Example 4.2 is equal to 0.023). However, the gauge reproducibility estimate in this analysis is more than three times greater than the result that we found in Example 4.2 ($\hat{\sigma}_{reproducibility}$ in Example 4.2 is equal to 0.012). This is because; there is a significant interaction between Operators and Parts in this experiment. As we can see, if there is interaction in the experiment, an experimental design approach is a better method for us to use to analyze the gauge measurement data.

(2) Mixed Effects Model

Now, we consider that the operators are a fixed factor, such as would occur if the operators were the only operators in the organization. We still consider that the parts are selected randomly from a larger population.

For this mixed model, we are using the same linear representation for the statistical model as for the random effects model (Equation (4.6)). Now however, O_i is a fixed effect, P_j is a random effect, and interaction $(OP)_{ij}$ can be treated as random effect. Since, O_i is a fixed effect we require that $\sum_{i=1}^{o} O_i = 0$, P_j is normally distributed with mean zero and variance σ_P^2, and $(OP)_{ij}$ is normally distributed with mean zero and variance $[(o-1)/o]\sigma_{OP}^2$.

The calculations of the sums of squares $SS_{Operator}$, SS_{Part}, $SS_{Interaction}$, SS_R, and SS_T are the same as for the random effects model. However, the expected mean squares and test statistics are slightly different as compared with the random effects model. The expected mean squares for each component are as follows:

$$E(MS_O) = \sigma_R^2 + n\sigma_{OP}^2 + \left(\frac{pn\sum_{i=1}^{o} O_i^2}{o-1} \right) \quad (4.32)$$

$$E(MS_P) = \sigma_R^2 + on\sigma_P^2 \quad (4.33)$$

$$E(MS_{OP}) = \sigma_R^2 + n\sigma_{OP}^2 \quad (4.34)$$

and

$$E(MS_R) = \sigma_R^2 \quad (4.35)$$

We can apply the same procedure that we described in the previous section on the *random effects model* to determine the test statistics for mixed effects model. Based on the expected mean squares, the statistics for testing $H_0 : O_i = 0$ is

$$F_0 = \frac{MS_O}{MS_{OP}} \tag{4.36}$$

where under the assumption of normality the ratio F_0 is distributed as $F_{(o-1),(o-1)(p-1)}$.

The test statistics for testing $H_0 : \sigma_P^2 = 0$ is as before

$$F_0 = \frac{MS_P}{MS_R} \tag{4.37}$$

with the ratio F_0 distributed as $F_{(p-1),op(n-1)}$.

The test statistics for testing $H_0 : \sigma_{OP}^2 = 0$ is

$$F_0 = \frac{MS_{OP}}{MS_R} \tag{4.38}$$

and the ratio F_0 is distributed as $F_{(o-1)(p-1),op(n-1)}$.

For this mixed model, the estimates of the overall mean and the fixed factor effects are as follows:

$$\hat{\mu} = \overline{X}_{...} \tag{4.39}$$

$$\hat{O}_i = \overline{X}_{i..} - \overline{X}_{...} \qquad i = 1, 2, \cdots, o \tag{4.40}$$

Also, the variance components of the random factor effects can be estimated as follows:

$$\hat{\sigma}_P^2 = \frac{MS_P - MS_R}{on} \tag{4.41}$$

$$\hat{\sigma}^2_{OP} = \frac{MS_{OP} - MS_R}{n} \tag{4.42}$$

and

$$\hat{\sigma}^2_R = MS_R \tag{4.43}$$

The analysis of variance for the gauge capability study (mixed model) is presented in Table 4.5. The estimates of variance components for gauge repeatability and reproducibility are found as with random effect model.

SOURCE OF VARIABILITY	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F-Ratio
Operators	$SS_{Operator}$	$o-1$	$MS_O = \frac{SS_{Operator}}{o-1}$	$F_0 = \frac{MS_O}{MS_{OP}}$
Parts	SS_{Part}	$p-1$	$MS_P = \frac{SS_{Part}}{p-1}$	$F_0 = \frac{MS_P}{MS_R}$
Operator by Part	$SS_{Interaction}$	$(o-1)(p-1)$	$MS_{OP} = \frac{SS_{Interaction}}{(o-1)(p-1)}$	$F_0 = \frac{MS_{OP}}{MS_R}$
Repeatability	SS_R	$op(n-1)$	$MS_R = \frac{SS_R}{op(n-1)}$	
Total	SS_T	$opn-1$		

Table 4.5 Analysis of variance for the two-factor mixed model

4-6 A Modification of the Classical Reproducibility Estimate

In Section 4-3, we discussed the classical reproducibility estimates for a gauge capability study. However, Montgomery and Runger (1993a) pointed out that the gauge reproducibility in a classical gauge study would be underestimated when the Operator by

Part interaction is large. The example we consider previously illustrates this point, where we found the estimate of gauge reproducibility from the classical R&R study to be $\hat{\sigma}'_{reproducibility} = 0.012$ where the analysis of variance estimate was much larger at $\hat{\sigma}_{reproducibility} = 0.0387$. We now consider why the classical reproducibility estimate produces such an underestimate in this situation. The mean value for operator i of the model in Equation (4.6) is

$$\overline{X}_{i..} = \mu + O_i + \frac{\sum_{j=1}^{p} P_j}{p} + \frac{\sum_{j=1}^{p} (OP)_{ij}}{p} + \frac{\sum_{j=1}^{p}\sum_{k=1}^{n} R_{(ij)k}}{pn} \tag{4.44}$$

We also saw that the variance estimate of gauge reproducibility in the classical gauge analysis is $\hat{\sigma}'_{reproducibility} = R_{\overline{\overline{X}}} / d_2$ (where the d_2 factor corresponds to the number of operators used in the study). From the results in Montgomery and Runger (1993a), the expected value of this reproducibility standard deviation is

$$E(\hat{\sigma}_{reproducibility}) = E\left(\frac{R_{\overline{X}_{i..}}}{d_2}\right) = \left(\sigma_O^2 + \frac{\sigma_{OP}^2}{p} + \frac{\sigma_R^2}{pn}\right)^{1/2} \tag{4.45}$$

where $R_{\overline{X}_{i..}} = \max_i(\overline{X}_{i..}, i = 1, 2, \cdots, o) - \min_i(\overline{X}_{i..}, i = 1, 2, \cdots, o)$, and the notation of $R_{\overline{X}_{i..}}$ and $R_{\overline{\overline{X}}}$ are the same as previously.

If we compare Equation (4.45) with Equation (4.29) ($\sigma^2_{reproducibility} = \sigma_O^2 + \sigma_{OP}^2$), the estimator in Equation (4.45) is biased. Obviously, the bias is due to the terms involving σ_{OP}^2 and σ_R^2 in Equation (4.45). Usually, it can be anticipated that the contribution to the bias from the term σ_R^2 is small. This is because the variation due to replications is not much larger than the variation due to operators, and the divisor pn from the number of parts and replications used in a gauge study will make the contribution from this term small.

However, the term σ_{OP}^2 can be an important component in the estimate of reproducibility in the gauge capability analysis. Normally, the number of parts used in the study is often large (25 parts in Example 4.2). For this reason, the gauge reproducibility in the classical gauge analysis could be seriously underestimated.

In spite of the problem of underestimation, the classical R&R analysis is worth using because the method is easy and simple. Montgomery and Runger (1993a) discussed an alternative method of estimation for reproducibility to correct the underestimation problem. The following method applied to the classical procedure will approximate the result of the ANOVA analysis when the Operator by Part interaction is large.

The procedure of this alternative method is as follows:

(1) Calculate $\overline{X}_{ij.}$, the average of the replicate measurements make by each operator on each part, for $i = 1, 2, \cdots, o$ and $j = 1, 2, \cdots, p$.

(2) For each part use the averages $\overline{X}_{ij.}$ to compute the range across the operators and denote this as R_j.

(3) Calculate $\overline{R}$ by averaging the ranges R_j over all parts.

(4) Divide $\overline{R}$ by the d_2 factor (the d_2 factor corresponding to the number of operators used in the study) to obtain $\hat{\sigma}^*_{reproducibility}$.

The expected value of this reproducibility estimator, $\hat{\sigma}^*_{reproducibility}$, is given by Montgomery and Runger (1993b):

$$E(\hat{\sigma}^*_{reproducibility}) = \left(\sigma_O^2 + \sigma_{OP}^2 + \frac{\sigma_R^2}{n} \right)^{1/2} \tag{4.46}$$

If we compare Equation (4.46) with Equation (4.29) ($\sigma^2_{reproducibility} = \sigma_O^2 + \sigma_{OP}^2$), the bias in Equation (4.46) is only due to the term involving σ_R^2. We can anticipate that if the variance component σ_R^2 is small relative to the variance component σ_O^2 and σ_{OP}^2, or if the number of replicates, n, in the study is large, then the reproducibility estimator in Equation (4.46) should be a reasonable estimator to use for estimating the reproducibility in the gauge capability study.

Example 4.6

In this example, we would like to use the above method to compute the gauge reproducibility estimate, $\hat{\sigma}^*_{reproducibility}$. We now reconsider the gauge measurement data in Table 4.2. In order to obtain the reproducibility estimate $\hat{\sigma}^*_{reproducibility}$, we have to calculate the averages, $\overline{X}_{ij.}$, and the ranges, R_j. For example, the range over operators for part 1 is $10.235 - 10.165 = 0.070$, for part 2 is $9.810 - 9.780 = 0.030$, and so on. The averages, $\overline{X}_{ij.}$, and the ranges, R_j, for all 25 parts are displayed in Table 4.6.

Part	Operator 1			Operator 2			Operator 3			
	Measurements			Measurements			Measurements			
	1	2	$\overline{x}_{1j\cdot}$	1	2	$\overline{x}_{2j\cdot}$	1	2	$\overline{x}_{3j\cdot}$	R_j
1	10.210	10.260	10.235	10.150	10.190	10.170	10.160	10.170	10.165	0.070
2	9.830	9.790	9.810	9.800	9.760	9.780	9.770	9.800	9.785	0.030
3	10.050	10.060	10.055	9.920	9.920	9.920	10.100	10.090	10.095	0.175
4	10.010	10.020	10.015	10.000	9.970	9.985	9.980	9.970	9.975	0.040
5	9.800	9.850	9.825	9.840	9.820	9.830	9.780	9.790	9.785	0.045
6	10.120	10.170	10.145	10.120	10.100	10.110	10.100	10.090	10.095	0.050
7	10.230	10.260	10.245	10.260	10.230	10.245	10.240	10.290	10.265	0.020
8	10.190	10.180	10.185	10.130	10.130	10.130	10.160	10.150	10.155	0.055
9	10.000	10.040	10.020	9.980	10.020	10.000	10.060	10.020	10.040	0.040
10	9.710	9.700	9.705	9.680	9.720	9.700	9.730	9.740	9.735	0.035
11	10.240	10.280	10.260	10.100	10.140	10.120	10.200	10.220	10.210	0.140
12	10.060	10.010	10.035	10.070	10.080	10.075	10.110	10.120	10.115	0.080
13	10.020	10.060	10.040	10.020	10.030	10.025	10.080	10.090	10.085	0.060
14	9.870	9.890	9.880	10.030	10.050	10.040	9.870	9.870	9.870	0.170
15	9.830	9.920	9.875	9.950	9.950	9.950	9.880	9.940	9.910	0.075
16	10.360	10.330	10.345	10.270	10.300	10.285	10.350	10.340	10.345	0.060
17	10.150	10.100	10.125	10.160	10.180	10.170	10.260	10.210	10.235	0.110
18	9.840	9.860	9.850	9.850	9.830	9.840	9.860	9.890	9.875	0.035
19	10.580	10.590	10.585	10.500	10.470	10.485	10.470	10.500	10.485	0.100
20	9.980	10.020	10.000	10.040	10.000	10.020	9.990	10.020	10.005	0.020
21	10.030	10.020	10.025	9.990	9.970	9.980	9.970	9.970	9.970	0.055
22	9.700	9.660	9.680	9.760	9.700	9.730	9.780	9.760	9.770	0.090
23	10.150	10.140	10.145	10.090	10.050	10.070	10.080	10.130	10.105	0.075
24	9.980	10.000	9.990	9.980	9.980	9.980	9.950	9.980	9.965	0.025
25	10.220	10.210	10.215	10.150	10.190	10.170	10.170	10.160	10.165	0.050
										$\overline{R} = 0.068$

Table 4.6 Data for Example 4.6

In Table 4.6, the average range, $\overline{R}$, is 0.068, and hence

$$\hat{\sigma}^{*}_{reproducibility} = \frac{\overline{R}}{d_2} = \frac{0.068}{1.693}$$
$$= 0.040$$

where the d_2 value is obtained from Table I in the Appendix with $n = 3$ reflecting the fact that three operators are involved in this experiment.

As we can see, the gauge reproducibility estimate obtained above ($\hat{\sigma}^*_{reproducibility} = 0.040$) is very close as compared to the result of the analysis of variance estimate ($\hat{\sigma}_{reproducibility} = 0.0387$).

4-7 Alternative Method for Negative Variance Estimates

One of the problems that may be encountered when the analysis of variance approach is used for estimating gauge repeatability and reproducibility is the estimate of gauge reproducibility may be negative. This occurs because of the need to include the interaction effect in the estimate of reproducibility.

We recall that the estimator of the variance component for the interaction between Operators and Parts in the Montgomery and Runger (1993a) model is

$$\hat{\sigma}^2_{OP} = \frac{MS_{OP} - MS_R}{n}$$

In gauge capability studies, it may often happen that the mean square for the Operator by Part interaction will be small. In such cases, the variance component estimate $\hat{\sigma}^2_{OP}$ may be negative. If this were to occur it is possible that the estimate of $\sigma^2_{reproducibility}$ which by Equation (4.29) is

$$\hat{\sigma}^2_{reproducibility} = \hat{\sigma}^2_O + \hat{\sigma}^2_{OP}$$

could also be negative.

To overcome this problem, should it arise, a procedure has been proposed by Milliken and Johnson (1984). They suggest that we begin by fitting the full model as previously outlined and test hypotheses concerning the various sources of variability. They also suggest that these tests be carried out at a level of significance α set at a value such as $\alpha < 0.3$. If all the components are significant at the chosen level, the variance components will be estimated. In this case, all the estimates should be found to be non-negative.

If one or more of the components are tested and found to be not significant at the chosen α level, the least important factors are eliminated from the model and the reduced model is then fitted. The analysis of variance tests are conducted on the components of the reduced model and again if one or more tests are not significant, the least important factors are eliminated and the model refitted. This is repeated until all components of the sources of variability are significantly different from zero. The variance components are then estimated and the repeatability and reproducibility estimates determined. These estimates will be non-negative.

Chapter 5

Confidence Interval Estimation of Components of Variance in Gauge Capability Studies

5-1 Confidence Intervals

In Chapter 4, we described how to use different methods to estimate the variance components of gauge capability studies. In this section, we will discuss how to construct a confidence interval on these variance components. If the normality assumptions are satisfied for the model given in Equation (4.6), then confidence intervals on the variance components of gauge capability studies can be obtained easily. The resulting confidence intervals can also provide insight for sample size determinations for designing gauge capability studies.

We first will consider the construction of a $100(1-\alpha)\%$ confidence interval on $\sigma^2_{repeatability}$. In order to construct a confidence interval on $\sigma^2_{repeatability}$, we can simply use the fundamental idea of a confidence interval for σ^2 based on a sample variance S^2. This can be represented as follows:

$$\frac{(n-1)S^2}{\chi^2_{\alpha/2,(n-1)}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,(n-1)}} \tag{5.1}$$

where S^2 is the estimator of σ^2, and $(n-1)$ is the degrees of freedom of the statistic S^2. Also, $\chi^2_{\alpha,\nu}$ is the value of the chi-square distribution with ν degrees of freedom, such that $P(\chi^2 > \chi^2_{\alpha,\nu}) = \alpha$.

From the model for the gauge capability study described by Montgomery and Runger (1993a), we can directly obtain the unbiased estimator for $\sigma^2_{repeatability}$. The estimator, $\hat{\sigma}^2_{repeatability}$, is obtained from the repeatability mean square, which in this study has $op(n-1)$ degrees of freedom.

The confidence interval on $\sigma^2_{repeatability}$ is then

$$\frac{op(n-1)\hat{\sigma}^2_{repeatability}}{\chi^2_{\alpha/2,op(n-1)}} \le \sigma^2_{repeatability} \le \frac{op(n-1)\hat{\sigma}^2_{repeatability}}{\chi^2_{1-\alpha/2,op(n-1)}} \tag{5.2}$$

where $\chi^2_{\alpha,op(n-1)}$ is the tabulated chi-square distribution value with $op(n-1)$ degrees of freedom with pre-selected α.

However, the exact confidence interval for $\sigma^2_{reproducibility}$ may not be found, because as noted previously there is no direct estimator of $\sigma^2_{reproducibility}$. There are no mean squares from the ANOVA that have expected value equal to $\sigma^2_O + \sigma^2_{OP}$. We saw that the estimator of $\sigma^2_{reproducibility}$ is based on two variance component estimators, one is the variance component estimator of Operator, σ^2_O, and the other is the variance component estimator of the Operator by Part interaction, σ^2_{OP}. Let us now recall the estimates of the variance components for σ^2_O and σ^2_{OP} in Equation (4.24) and Equation (4.26)

$$\hat{\sigma}^2_O = \frac{MS_O - MS_{OP}}{pn} \quad \text{and} \quad \hat{\sigma}^2_{OP} = \frac{MS_{OP} - MS_R}{n}$$

Because of the nature of these estimators and hence the estimator for $\sigma^2_{reproducibility}$ and σ^2_{gauge} derived from them, exact confidence intervals cannot be determined. However, Montgomery and Runger (1993b) provided details for the construction of approximate $100(1-\alpha)\%$ confidence interval estimates for $\sigma^2_{reproducibility}$ and σ^2_{gauge}. An approximate $100(1-\alpha)\%$ confidence interval on $\sigma^2_{reproducibility}$ is given by

$$\frac{v\hat{\sigma}^2_{reproducibility}}{\chi^2_{\alpha/2,v}} \le \sigma^2_{reproducibility} \le \frac{v\hat{\sigma}^2_{reproducibility}}{\chi^2_{1-\alpha/2,v}} \tag{5.3}$$

where

$$v = \frac{(\hat{\sigma}^2_{reproducibility})^2}{\dfrac{(1/np)^2 MS_O^2}{o-1} + \dfrac{[(p-1)/np]^2 MS_{OP}^2}{(o-1)(p-1)} + \dfrac{(1/n)^2 MS_R^2}{op(n-1)}} \tag{5.4}$$

It is easy to see why it may not be possible to obtain the exact confidence interval on σ^2_{gauge}. The estimator of σ^2_{gauge} is also based on two variance component estimators, one is the estimator of repeatability, $\sigma^2_{repeatability}$, and the other is the estimator of reproducibility, $\sigma^2_{reproducibility}$. Since the estimator of σ^2_{gauge} relies upon the estimator of $\sigma^2_{reproducibility}$, we face the same problem as for the construction of the confidence interval on $\sigma^2_{reproducibility}$. Montgomery and Runger (1993b) also provided an approximate $100(1-\alpha)\%$ confidence interval on σ^2_{gauge} as

$$\frac{u\hat{\sigma}^2_{gauge}}{\chi^2_{\alpha/2,u}} \le \sigma^2_{gauge} \le \frac{u\hat{\sigma}^2_{gauge}}{\chi^2_{1-\alpha/2,u}} \tag{5.5}$$

where

$$u = \frac{(\hat{\sigma}^2_{gauge})^2}{\frac{(1/np)^2 MS_O^2}{o-1} + \frac{[(p-1)/np]^2 MS_{OP}^2}{(o-1)(p-1)} + \frac{[(n-1)/n]^2 MS_R^2}{op(n-1)}} \tag{5.6}$$

Example 5.1

In this example, we will construct the confidence intervals on $\sigma^2_{repeatability}$, $\sigma^2_{reproducibility}$ and σ^2_{gauge} using the data from Table 4.2. The variance component estimates for $\sigma^2_{repeatability}$ and $\sigma^2_{reproducibility}$ using the data were obtained in Example 4.5; the estimates for these three variance components are found to be

$$\hat{\sigma}^2_{repeatability} = 0.0005$$

$$\hat{\sigma}^2_{reproducibility} = 0.0015$$

and hence

$$\hat{\sigma}^2_{gauge} = 0.0005 + 0.0015 = 0.0020$$

We will now construct the 95% confidence intervals for $\sigma^2_{repeatability}$, $\sigma^2_{reproducibility}$ and σ^2_{gauge}. Using Equation (5.2), the 95% confidence interval on $\sigma^2_{repeatability}$ can be computed as follows:

$$\frac{op(n-1)\hat{\sigma}^2_{repeatability}}{\chi^2_{\alpha/2,op(n-1)}} \le \sigma^2_{repeatability} \le \frac{op(n-1)\hat{\sigma}^2_{repeatability}}{\chi^2_{1-\alpha/2,op(n-1)}}$$

$$\frac{75(0.0005)}{100.8390} \le \sigma^2_{repeatability} \le \frac{75(0.00050)}{52.9419}$$

$$0.0004 \le \sigma^2_{repeatability} \le 0.0007$$

Using Equations (5.3) and (5.4), an approximate 95% confidence interval on $\sigma^2_{reproducibility}$ can be computed as follows:

$$\frac{\nu\hat{\sigma}^2_{reproducibility}}{\chi^2_{\alpha/2,\nu}} \le \sigma^2_{reproducibility} \le \frac{\nu\hat{\sigma}^2_{reproducibility}}{\chi^2_{1-\alpha/2,\nu}}$$

where

$$\nu = \frac{(\hat{\sigma}^2_{reproducibility})^2}{\frac{(1/np)^2 MS_O^2}{o-1} + \frac{[(p-1)/np]^2 MS_{OP}^2}{(o-1)(p-1)} + \frac{(1/n)^2 MS_R^2}{op(n-1)}}$$

$$= \frac{(0.0015)^2}{\frac{(1/50)^2(0.0055)^2}{2} + \frac{(24/50)^2(0.0035)^2}{48} + \frac{(1/2)^2(0.0005)^2}{75}}$$

$$= 34.2553 \approx 34$$

And hence, an approximate 95% confidence interval on $\sigma^2_{reproducibility}$ is

$$\frac{34(0.0015)}{51.9660} \le \sigma^2_{reproducibility} \le \frac{34(0.0015)}{19.8062}$$

$$0.0010 \le \sigma^2_{reproducibility} \le 0.0026$$

We can also obtain an approximate 95% confidence interval on σ^2_{gauge} using Equations (5.5) and (5.6), and this can be computed as follows:

$$\frac{u\hat{\sigma}^2_{gauge}}{\chi^2_{\alpha/2,u}} \le \sigma^2_{gauge} \le \frac{u\hat{\sigma}^2_{gauge}}{\chi^2_{1-\alpha/2,u}}$$

where

$$\begin{aligned}
u &= \frac{(\hat{\sigma}^2_{gauge})^2}{\frac{(1/np)^2 MS_O^2}{o-1} + \frac{[(p-1)/np]^2 MS_{OP}^2}{(o-1)(p-1)} + \frac{[(n-1)/n]^2 MS_R^2}{op(n-1)}} \\
&= \frac{(0.0020)^2}{\frac{(1/50)^2 (0.0055)^2}{2} + \frac{(24/50)^2 (0.0035)^2}{48} + \frac{(1/2)^2 (0.0005)^2}{75}} \\
&= 60.8982 \approx 61
\end{aligned}$$

Hence, an approximate 95% confidence interval on σ^2_{gauge} is

$$\begin{gathered}
\frac{61(0.0020)}{84.4764} \le \sigma^2_{gauge} \le \frac{61(0.0020)}{41.3031} \\
0.0014 \le \sigma^2_{gauge} \le 0.0030
\end{gathered}$$

We can use Equations (5.5) and (5.6) to develop a confidence interval on the P/T ratio. The P/T ratio is equal to $6\sigma/(USL - LSL)$. An approximate $100(1-\alpha)\%$ confidence interval for P/T ratio is then

$$\frac{6\sqrt{\frac{u\hat{\sigma}^2_{gauge}}{\chi^2_{\alpha/2,u}}}}{USL - LSL} \le \frac{6\sigma_{gauge}}{USL - LSL} \le \frac{6\sqrt{\frac{u\hat{\sigma}^2_{gauge}}{\chi^2_{1-\alpha/2,u}}}}{USL - LSL} \tag{5.7}$$

The approximate confidence intervals (5.3), (5.5) and (5.7) are appropriate, if the estimates of variance component from the full factorial model are nonnegative. In Section 4-7, we described the modified method that can be used when negative estimates of variance component are encountered. Milliken and Johnson (1984) also provided the approximate confidence intervals for this modified method.

5-2 Sample Size Determination

As we mentioned earlier, the confidence intervals we obtained can provide some insight for sample size determination. In this section, we will discuss the effect of the number of replications, parts and operators on the sensitivity of the confidence intervals.

We now look at the confidence intervals from Example 5.1 and examine what would happen to those confidence intervals if n, p and o were changed. First of all, we would like to consider the confidence interval for $\sigma^2_{repeatability}$ in Equation (5.2). The precision of the confidence interval for $\sigma^2_{repeatability}$ can be determined by the following ratios

$$\frac{op(n-1)}{\chi^2_{\alpha/2,\,op(n-1)}} \quad \text{and} \quad \frac{op(n-1)}{\chi^2_{1-\alpha/2,\,op(n-1)}} \tag{5.8}$$

In Example 5.1, the 95% confidence interval for $\sigma^2_{repeatability}$ can also be expressed as

$$\frac{op(n-1)}{\chi^2_{0.025,\,op(n-1)}}\hat{\sigma}^2_{repeatability} \le \sigma^2_{repeatability} \le \frac{op(n-1)}{\chi^2_{0.975,\,op(n-1)}}\hat{\sigma}^2_{repeatability}$$

$$\frac{75}{100.8390}\hat{\sigma}^2_{repeatability} \le \sigma^2_{repeatability} \le \frac{75}{52.9419}\hat{\sigma}^2_{repeatability}$$
$$0.74\hat{\sigma}^2_{repeatability} \le \sigma^2_{repeatability} \le 1.42\hat{\sigma}^2_{repeatability}$$

We can obtain a measure of the precision of the confidence by considering the width of the interval. If we divide by $\hat{\sigma}^2_{repeatability}$ in the above confidence interval, the width of the interval can be expressed as $(1.42 - 0.74) = 0.68$. In other word, the width of this interval is 68% of $\hat{\sigma}^2_{repeatability}$

In order to have more precise variance component estimates, it is very important to consider which parameters (o, p or n) might be changed in the experiment. From the result above, the width of the confidence interval would be reduced if one of the parameters, such as the number of replications, the number of parts or the number of operators were increased. However, it is necessary for us to consider which of these parameters should be changed. In our example, $p = 25$, $o = 3$, and $n = 2$, and hence $op(n-1) = 75$. We can double the value of $op(n-1)$ by doubling either the number of parts or operators. As a result, the width of the confidence interval would become 46% of $\hat{\sigma}^2_{repeatability}$ ($0.81\hat{\sigma}^2_{repeatability}$ to $1.27\hat{\sigma}^2_{repeatability}$). However, if we were to double the number of replications from $n = 2$ to $n = 4$, then the value of $op(n-1)$ will increase to 225. In this case, the width of the confidence interval becomes 37% of $\hat{\sigma}^2_{repeatability}$ ($0.84\hat{\sigma}^2_{repeatability}$ to $1.21\hat{\sigma}^2_{repeatability}$). As we can see, the width of the confidence interval for $\sigma^2_{repeatability}$ will reduce as the value of $op(n-1)$ increases.

For considering the precision of the confidence intervals for $\sigma^2_{reproducibility}$ and σ^2_{gauge}, we can apply the same idea that we described above. The width of the confidence intervals for $\sigma^2_{reproducibility}$ and σ^2_{gauge} are determined by these ratios

$$\frac{v}{\chi^2_{\alpha/2,v}} \quad \text{and} \quad \frac{v}{\chi^2_{1-\alpha/2,v}} \tag{5.9}$$

and

$$\frac{u}{\chi^2_{\alpha/2,u}} \quad \text{and} \quad \frac{u}{\chi^2_{1-\alpha/2,u}} \tag{5.10}$$

where the values v and u are the degrees of freedom given in Equations (5.3) and Equation (5.5).

As we can see, the width of the confidence intervals for $\sigma^2_{reproducibility}$ and σ^2_{gauge} will be reduced if the values of v and u in Equations (5.9) and (5.10) increase. Also, there will be a greater impact on the width for both confidence intervals ($\sigma^2_{reproducibility}$ and σ^2_{gauge}) through increasing n rather than for p or o. For example, if we double the number of operators in Example 4.2 from $o=3$ to $o=6$, then the degrees of freedom v is approximately equal to 85. The width of the confidence interval is 62% of $\hat{\sigma}^2_{reproducibility}$ ($0.76\hat{\sigma}^2_{reproducibility}$ to $1.38\hat{\sigma}^2_{reproducibility}$). However, by doubling the number of replications from $n=2$ to $n=4$, the degrees of freedom v is approximately equal to 138. The width of the confidence interval is then 49% of $\hat{\sigma}^2_{reproducibility}$ ($0.80\hat{\sigma}^2_{reproducibility}$ to $1.29\hat{\sigma}^2_{reproducibility}$).

As we can see increasing any of the components of the study, o, p or n, will have the effect of decreasing the width of the confidence interval, or increasing the precision of the estimates. It may not be practical to have additional operators and increasing the number of parts. However, we note that increasing the number of replicate measurements can achieve a substantial reduction in the width of the interval.

Chapter 6

Assessing Additional Sources of Gauge Measurement Error

In the previous two sections, we have concentrated our discussion on the variability associated with gauge repeatability and gauge reproducibility. However, there are many additional sources of error that would affect the results of measurement. In this section, we will discuss three other sources of measurement error - gauge accuracy, linearity and stability.

6-1 Gauge Accuracy

In Section 2-1, we defined *Accuracy* as the closeness of agreement between a test result and the accepted reference value. We also have described the concept of gauge accuracy in Section 2-2. Here, we will consider how to apply a test procedure to determine the gauge accuracy. In order to examine gauge accuracy, operators take p parts and measure them. Given the definition of accuracy, we would like to be able to calculate the difference between the test result and the accepted reference value of the part. Let T_i denote the accepted reference value of the i^{th} part that may have been obtained by using a highly sensitive and more accurate measuring instrument. The data and differences can be displayed as follows:

Part	Measurements	Difference
1	X_1	$D_1 = X_1 - T_1$
2	X_2	$D_2 = X_2 - T_2$
•	•	•
•	•	•
p	X_p	$D_p = X_p - T_p$
Mean	$\overline{X}$	$\overline{D}$

Table 6.1

Kane (1989) described a simple test procedure for examining the gauge accuracy based upon the well-known t test procedure. He suggested that we could use the mean of the differences, $\overline{D}$, in Table 6.1 to calculate a test statistic. The test statistic is as follows:

$$t = \frac{|\overline{D}|}{S_D / \sqrt{p}} \tag{6.1}$$

where

$$S_D = \sqrt{\frac{(D_1 - \overline{D})^2 + (D_2 - \overline{D})^2 + \cdots + (D_p - \overline{D})^2}{p-1}} \tag{6.2}$$

and p is the number of parts that have been selected. If the selected gauge is accurate, then we would expect the mean difference to be close to zero.

If the test statistics t is larger than the $t_{\alpha/2,(p-1)}$ (Student's t distribution with $p-1$ degrees of freedom and pre-selected α), then we can conclude that there is evidence of a significant difference between the measurement value obtained using the selected gauge and the accepted reference value determined from a high accuracy and sensitive gauge.

6-2 Gauge Linearity

In Section 6-1, we provided a discussion on gauge accuracy. An additional concern with respect to the accuracy of the gauge is how well the gauge performs across the operating range of the instrument. This characteristic of the instrument is referred to as gauge linearity, which is defined as follows:

Linearity: the change in accuracy through the operating range of the gauge.

For a given gauge we can graph the measured values over the operating range to examine the linearity of a gauge. In Figure 6.1, the graph shows a gauge that displays a lack of linearity.

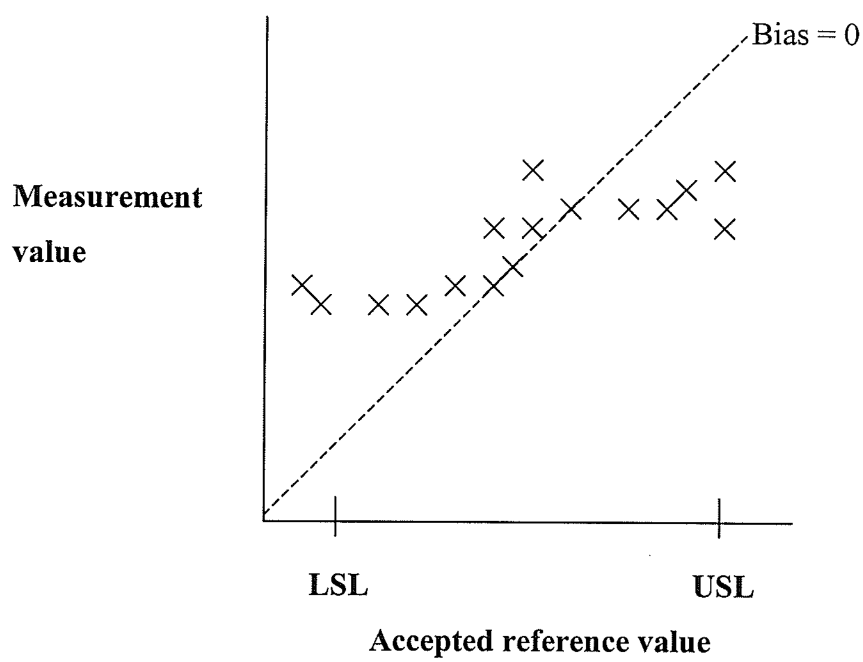


Figure 6.1 Example of a gauge lack of linearity

In above Figure, the gauge is accurate around the midpoint of the specification limits. However, we can see larger bias as we approach the extremes of the specification interval.

Usually, the problem of lack of gauge linearity can be minimized. One possible way is to ensure that parts near the upper and lower end of the specification limits have been examined before the gauge is purchased. Moreover, it is then necessary to check the gauge periodically thereafter. Most often, an instrument has good accuracy at the centre of the specification limits. However, it might not be accurate at the lower or upper end of the specification limit.

6-3 Gauge Stability

An important consideration for a gauge is for the measurements that are made by a particular gauge to remain stable over a period of time. One way to examine the stability of the gauging process is to evaluate and to monitor the process using control charts. The control chart technique is valuable in identifying any special causes of variation that may be present in the process. To assess gauge stability, we would use the same gauge to obtain measurements on the same parts repeatedly over time. Kane (1989) described three types of control charts to assess gauge stability. We also discuss the use of a cumulative-sum (or Cusum) control chart to detect small shifts in the measuring system.

- Control Charts

Kane (1989) described the following three types of control charts:

(1) Gauge Target Control Chart

To construct a Gauge Target Control Chart, we will make ongoing repeated measurements of the same p (e.g. $p = 5$) parts. The procedure is as follows:

Step 1:

(1) A collection of p parts is selected that span the specification limits.

(2) Using a sensitive and accurate gauge we measure these p parts and designate these measured values as the accepted reference values for these p parts.

(3) Let $T_1, T_2, \cdots\cdots, T_p$ be the accepted reference values of these p parts. Their average is

$$\overline{T} = \frac{T_1 + T_2 + \cdots\cdots + T_p}{p} \tag{6.3}$$

This average, $\overline{T}$, is the target value for the average of repeated measurements to be made on this reference set of parts.

Step 2:

We will now conduct a gauge repeatability study on a large collection of parts, say $P = 30$. This study must be set up using strict repeatability conditions with a single operator and the gauge in good working condition. The purpose of this study is to obtain a reliable estimate of gauge repeatability, $\hat{\sigma}_{repeatability}$. The crucial concern here is ensuring that all measurements are made under repeatability conditions.

Using the estimate of repeatability, we can now construct the Gauge Target Control Chart as follows:

$$\begin{aligned} UCL &= \overline{T} + A\hat{\sigma}_{repeatability} \\ CL &= \overline{T} \\ LCL &= \overline{T} - A\hat{\sigma}_{repeatability} \end{aligned} \tag{6.4}$$

The value of the constant A, which is equal to $3/\sqrt{p}$, is obtained from the Table I in the Appendix.

At regular and specified intervals of time, the same reference set of p parts is re-measured and the mean measurement value, $\overline{X}$, is obtained. The mean measurement value at time t, $\overline{X}_t$, can be denoted as

$$\overline{X}_t = \frac{X_1 + X_2 + \cdots\cdots + X_p}{p} \qquad \text{for } t = 1, 2, \cdots\cdots \tag{6.5}$$

We plot the mean value on the control chart at each time interval. Any patterns or out of control points noted on the chart indicate a lack of stability for the gauge.

(2) Gauge Difference Control Chart

The Gauge Difference Control Chart is an alternative method to the Gauge Target Control Chart. The idea of this control chart is to use the differences between the individual gauge measurements, $X_1, X_2, \cdots\cdots, X_p$, and the accepted reference values, $T_1, T_2, \cdots\cdots, T_p$. The differences D_i can be denoted as

$$D_i = X_i - T_i \qquad \text{for } i = 1, 2, \cdots\cdots, p \tag{6.6}$$

The mean of these differences $\overline{D}$ and the corresponding standard deviation S_D can be calculated as follows:

$$\overline{D} = \frac{\sum_{i=1}^{p} D_i}{p} \quad \text{and} \quad S_D = \sqrt{\frac{\sum_{i=1}^{p} (D_i - \overline{D})^2}{p-1}} \tag{6.7}$$

The statistics $\overline{D}$ and S_D are the plotting statistics that will be used to monitor the gauge over time.

If the gauge is working as desired, we would expect the mean of the differences to be zero. As a result, we will set the centre line on the $\overline{X}$ control chart at 0.

The centre line and the control limits for the Gauge Difference $\overline{X}$ chart are given as follows:

$$\begin{aligned} UCL &= A\hat{\sigma}_{repeatability} \\ CL &= 0 \\ LCL &= -A\hat{\sigma}_{repeatability} \end{aligned} \tag{6.8}$$

The centre line and control limits for the Gauge Difference S chart are given by:

$$\begin{aligned} UCL &= B_6\hat{\sigma}_{repeatability} \\ CL &= c_4\hat{\sigma}_{repeatability} \\ LCL &= B_5\hat{\sigma}_{repeatability} \end{aligned} \tag{6.9}$$

The constants A, B_5, B_6 and c_4 are obtained from the Table I in the Appendix.

In Equations (6.8) and (6.9), the estimate of gauge repeatability $\hat{\sigma}_{repeatability}$ is to be obtained through the same procedure that we described in Step 2 for the Gauge Target Control Chart.

In order to maintain this control chart, we have to re-measure the same p parts and subtract the accepted reference values to obtain the differences periodically over time. Then we can use the Equations (6.7) to calculate the mean $\overline{D}$ and standard deviation S_D at each time interval and plot these statistics on the Gauge Difference $\overline{X}$ and S chart. Again, the charts would be examined for unusual patterns of variability that would signal a lack of gauge stability.

(3) Gauge Repeatability Control Chart

For the previous two control charts, it is necessary at each new time point to measure the same reference set of p parts and compare the measured value against the accepted reference value T_i. In the following method, it is not necessary to use the same p parts and compare them to the accepted reference value. This alternative approach is to focus on the differences between two readings of measurements on each part. If the measurements that are made by the gauge remain stable over a period of time, then the differences and the standard deviation of the differences between two readings will be small. The procedure to construct a Gauge Repeatability Control Chart is as follows:

(1) Select p parts and measure these p parts twice. Denoted the first reading as X_i and second reading as Y_i, $i = 1, 2, \cdots\cdots, p$.

(2) Calculate the differences between first reading X and second reading Y on the part, and denoted the differences D_i as

$$D_i = X_i - Y_i \qquad \text{for } i = 1, 2, \cdots\cdots, p \tag{6.10}$$

(3) Calculate the mean and standard deviation of these differences:

$$\overline{D} = \frac{\sum_{i=1}^{p} D_i}{p} \quad \text{and} \quad S_D = \sqrt{\frac{\sum_{i=1}^{p}(D_i - \overline{D})^2}{p-1}} \tag{6.11}$$

For an ongoing process, we have to re-select another p parts and measure them. We calculate the mean of the differences $\overline{D}_t$ and the standard deviation of the differences S_{D_t} at each time t using Equations (6.11). Then, we can plot this mean and standard deviation on the Gauge Repeatability $\overline{X}$ and S control chart.

The centre line and control limits for Gauge Repeatability $\overline{X}$ chart are as follows:

$$\begin{aligned} UCL &= \overline{\overline{D}} + A\sqrt{2}\,\hat{\sigma}_{repeatability} \\ CL &= \overline{\overline{D}} \\ LCL &= \overline{\overline{D}} - A\sqrt{2}\,\hat{\sigma}_{repeatability} \end{aligned} \tag{6.12}$$

where $\overline{\overline{D}}$ is the overall mean of subgroup $\overline{D}$ values. We would likely assume this overall mean as 0 and hence the centre line is often set equal to 0.

The S chart for gauge repeatability is as follows:

$$\begin{aligned} UCL &= B_6\sqrt{2}\hat{\sigma}_{repeatability} \\ CL &= c_4\sqrt{2}\hat{\sigma}_{repeatability} \\ LCL &= B_5\sqrt{2}\hat{\sigma}_{repeatability} \end{aligned} \tag{6.13}$$

The constants A, B_5, B_6 and c_4 are obtained from the Table I in the Appendix.

The estimate of gauge repeatability $\hat{\sigma}_{repeatability}$ in Equations (6.12) and (6.13) can be obtained through the same procedure that we described in Step 2 for the Gauge Target Control Chart.

Again, the charts would be examined for any unusual patterns or out of control points that would indicate a lack of gauge stability.

(4) Cumulative-Sum (or Cusum) Control Chart

The control charts that we described above have a significant limitation. Those control charts are based on the principles of the Shewhart control chart. Montgomery (2001b) pointed out that the Shewhart control chart is insensitive to detect small shifts in the process. The Shewhart control chart is effective if the magnitude of the shift is as large or larger than 1.5σ. For this reason, we might wish to use other control charts that can be more effective in detecting small shifts in the measuring process. The cumulative-sum (or cusum) control chart is one chart that is designed to detect small shifts in the process. We will discuss how to use a cusum control chart to monitor the stability of an instrument.

For checking the stability of the gauge, the first step is, as previously, to select p parts and measure these p parts using a sensitive and accurate gauge. These measured values, denoted as $T_1, T_2, \cdots\cdot, T_p$ can be considered as the accepted reference values of these p parts. The second step is to use another gauge (the gauge that we would like to check for gauge stability) to measure the same p parts and record the measurement values. We then can use these resulting measurements to construct a cusum control chart. In order to maintain the gauge stability, it is necessary to re-measure the same p parts on an ongoing basis and plot these measurement values on the constructed cusum chart. Out of control signals on the chart would indicated that the gauge lacks stability.

For constructing the cusum chart, we calculate the average of the deviations of the measured values from the corresponding accepted reference values at time j. This can be denote as

$$\begin{aligned}\overline{D}_j &= \frac{1}{p}\sum_{i=1}^{p}(X_{ij} - T_i) \\ &= \overline{X}_j - \overline{T}\end{aligned} \qquad \begin{cases} i = 1, 2, \cdots, p \\ j = 1, 2, \cdots\cdots \end{cases} \tag{6.14}$$

Then we calculate the cumulative sums of these averages. The quantity that we plot on the cumulative-sum control chart is

$$C_j = \overline{D}_j + C_{j-1} \qquad \text{for } j = 1, 2, \cdots\cdots \tag{6.15}$$

The starting value for the cusum, C_0, is assumed to be zero.

The following cusum chart that we will present is called a **Tabular Cusum Control Chart**. In order to construct this control chart, we need to calculate two tabular cusums, one-sided upper cusum, C^+, and one-sided lower cusum, C^-.

The one-sided upper and lower cusums can be computed as

$$C_j^+ = \max[0, \overline{X}_j - (\overline{T} + K) + C_{j-1}^+] \tag{6.16}$$

and

$$C_j^- = \max[0, (\overline{T} - K) - \overline{X}_j + C_{j-1}^-] \tag{6.17}$$

where C_0^+ and C_0^- are taken to be zero.

In Equations (6.16) and (6.17), the value K is called the **reference value** (or the **allowance** or the **slack value**). This can be defined as one-half the magnitude of the shift that we would like to detect. We can express this as

$$K = \frac{\delta\sigma_{gauge}}{2} \tag{6.18}$$

For example, if we want to detect shift of $1\sigma_{gauge}$, then the value K will be equal to $\sigma_{gauge}/2$.

If either C^+ or C^- exceed a **decision interval H,** then the measurement process can be considered to be out-of-control. The decision interval H is frequently taken to be five times the gauge standard deviation. Hence

$$H = 5\sigma_{gauge} \tag{6.19}$$

The centre line and control limits of the Tabular Cusum control chart are given as follows:

$$\begin{aligned} UCL &= H \\ CL &= 0 \\ LCL &= -H \end{aligned} \tag{6.20}$$

We can also use the cusum control chart to assist with determining when the shift was started to occur. A counter denoted by N^+ and N^- is use to record the number of the consecutive periods that C^+ and C^- are non-zero. Based on the signal that indicated an out-of-control situation and the corresponding counter value (N^+ or N^-), we can then determine when the process was last in control. For example, if the out-of-control signal is given at period 15 say, and the corresponding counter value is $N^+ = 6$. This indicated that the measurement process was last in control at period $15-6=9$. The shift has probably occurred between periods 9 and 10. This will assist in the determination of what may have happened to cause the gauge to shift, and what corrective action needs to be taken with respect to the gauge and the measurements made by it following the shift.

Chapter 7

Comparing Accuracy Between Two Measurement Instruments

7-1 Test for Difference in Precision Between Two Instruments

For any production process, it is very important that the measurements made on the same parts by vendor and customer inspection departments are in agreement. A disagreement between vendor and customer inspection departments in measurements might result, for example, in good product shipped by the vendor being rejected by the customer. For preventing such a problem, we have to ensure that similar operational procedures for obtaining the measurements are used by both the vendor and the customer inspection departments. In order to determine if there are discrepancies, we can use some simple procedures to test for any significant differences in accuracy across measuring instruments being used by the parties involved. In this section, we will introduce the procedures for testing precision and trueness between two instruments. Maloney and Rastogi (1970) discussed a significance test for the difference in precision between two instruments. Morgan (1939) was the first to develop this significance test.

We here consider testing a difference in precision between two instruments. To test for a difference between two instruments, we might consider using a test of the difference between the variances of the two instruments. In order to compare two instruments, each instrument measures the same set of p parts. The measured values for instrument 1 are denoted as

$$X_1, X_2, \cdots\cdot, X_p \tag{7.1}$$

while the values for instrument 2 are

$$Y_1, Y_2, \cdots\cdot, Y_p \tag{7.2}$$

Usually, the F test involving the ratio of the two sample variances would be used for comparing variances from two independent samples. However, Farnum (1994) pointed out that F test does not apply in this setting. This is because, each instrument has measured the same parts, and hence, the measurements are not independent. Morgan (1939) developed a significance test for this setting, and this test involves the usual test of significance associated with the correlation coefficient, r.

In order to construct this significance test, we first calculate sums and differences of these two measured values between instrument 1 and instrument 2 on each part. These are denoted as

$$S_i = X_i + Y_i \qquad i = 1, 2, \cdots\cdot, p \tag{7.3}$$

and

$$D_i = X_i - Y_i \qquad i = 1, 2, \cdots\cdot, p \tag{7.4}$$

so that

$$X_i = \frac{1}{2}(S_i + D_i) \qquad i = 1, 2, \cdots\cdot, p \tag{7.5}$$

and

$$Y_i = \frac{1}{2}(S_i - D_i) \qquad i = 1, 2, \cdots, p \tag{7.6}$$

The variances of X and Y can be expressed as

$$\sigma_X^2 = \frac{1}{4}(\sigma_S^2 + \sigma_D^2 + 2\rho_{SD}\sigma_S\sigma_D) \tag{7.7}$$

and

$$\sigma_Y^2 = \frac{1}{4}(\sigma_S^2 + \sigma_D^2 - 2\rho_{SD}\sigma_S\sigma_D), \tag{7.8}$$

respectively.

Based on Equations (7.7) and (7.8), we can see that the test for the hypothesis $H_0 : \rho_{SD} = 0$ is equivalent to the test for $H_0 : \sigma_X^2 = \sigma_Y^2$.

Maloney and Rastogi (1970) have shown that the test statistic can be derived using the correlation coefficient between the sums S_i and the differences D_i as

$$t = \frac{r\sqrt{p-2}}{\sqrt{1-r^2}} \tag{7.9}$$

where

$$r = \frac{\sum_{i=1}^{p}(S_i - \overline{S})(D_i - \overline{D})}{\sqrt{\sum_{i=1}^{p}(S_i - \overline{S})^2 \sum_{i=1}^{p}(D_i - \overline{D})^2}} \tag{7.10}$$

We can conclude that there is a difference in precision of two instruments at α level of significance, if $|t| > t_{\alpha/2}$ (where $t_{\alpha/2}$ is the Student's t distribution with $p-2$ degrees of freedom with pre-selected α).

7-2 Test for Difference in Trueness Between Two Instruments

Farnum (1994) provided a test procedure for testing for a difference in trueness between two instruments. The mean $\overline{D}$ and standard deviation S_D of the differences as given in Equation (7.4) are

$$\overline{D} = \frac{\sum_{i=1}^{p} D_i}{p} \tag{7.11}$$

and

$$S_D = \sqrt{\frac{\sum_{i=1}^{p} (D_i - \overline{D})^2}{p-1}} \tag{7.12}$$

The usual **matched pairs** test can be used to compare the mean of differences between two instruments. The test statistic then is computed as follows:

$$t = \frac{\overline{D}\sqrt{p}}{s_D} \tag{7.13}$$

We can conclude that there is a difference in trueness of two instruments at α level of significance, if $|t| > t_{\alpha/2}$ (where $t_{\alpha/2}$ is the Student's t distribution with $p-1$ degrees of freedom with pre-selected α).

Chapter 8

A Gauge Capability Study Using Semiconductor Data

In this section, we will consider an example using the experimental design approach for the analysis of a gauge capability study. This will include the analysis of variance for the nested and factorial design model, and confidence intervals on $\sigma^2_{repeatability}$, $\sigma^2_{reproducibility}$ and σ^2_{gauge}. The study is described in a paper by Borror, Montgomery and Runger (1997) and the authors have kindly provided a version of the data set for our use here. The study arises from a semiconductor manufacturing process. For this study, four specific sites are located on a semiconductor wafer and four measurements are recording at each of these sites on a wafer. For each day, data are collected from three shifts (operators). Further, this experiment was conducted over a seven-day period. In the data set, there are 336 individual observations. The data that we have displayed in Table 8.1 are collected from the third day. The complete data set is attached in Table II in the Appendix.

Day	Shift	Site	Measurements			
3	1	1	31.19	30.60	30.91	30.70
3	1	2	30.71	30.72	30.77	30.73
3	1	3	30.75	30.71	30.77	30.74
3	1	4	30.79	30.76	30.65	30.72
3	2	1	30.81	30.77	30.70	31.43
3	2	2	30.80	30.71	30.78	30.77
3	2	3	30.76	30.80	30.77	30.72
3	2	4	30.86	30.72	30.62	30.94
3	3	1	30.57	30.36	30.58	30.59
3	3	2	30.65	30.67	30.73	30.75
3	3	3	30.76	30.75	30.74	30.76
3	3	4	30.79	30.70	30.66	30.72

Table 8.1

Before conducting any data analysis, it is recommended that graphical displays of the data be made. First of all, we use a *box and whisker* plot to show the distribution of data on a day-by-day basis.

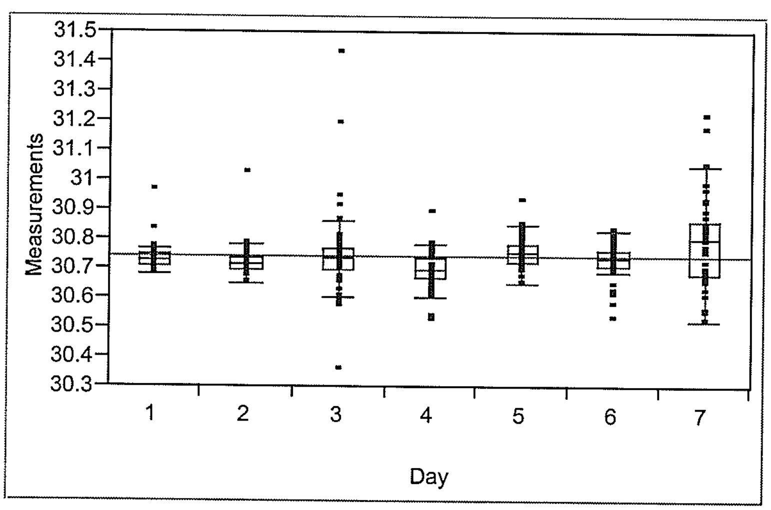


Figure 8.1 Box and whisker plot – All 7 days

In Figure 8.1, we can see that the median values for all 7 days are almost the same. However, the variation for Day 3 and Day 7 is larger than for the other 5 days. Also, some outliers have been noted for all seven days.

Now, we would like to determine why Day 3 and Day 7 have larger variation than the other 5 days. Again, we use side-by-side box and whisker plots to display the data for Day 3 and Day 7 across all three shifts.

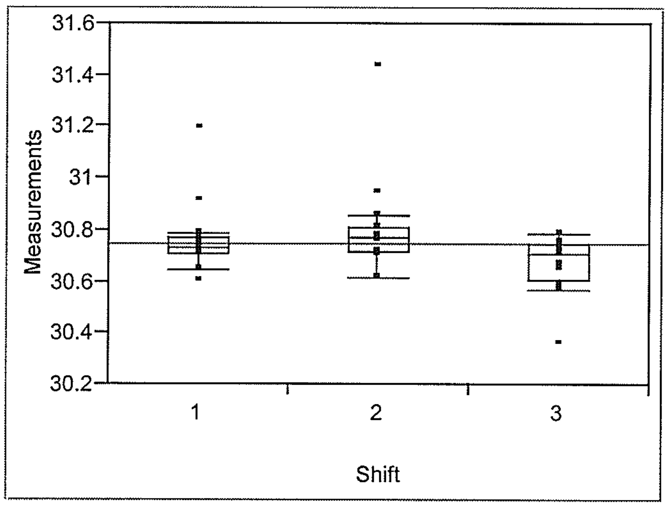


Figure 8.2 Box and whisker plot – Day 3 by Shift

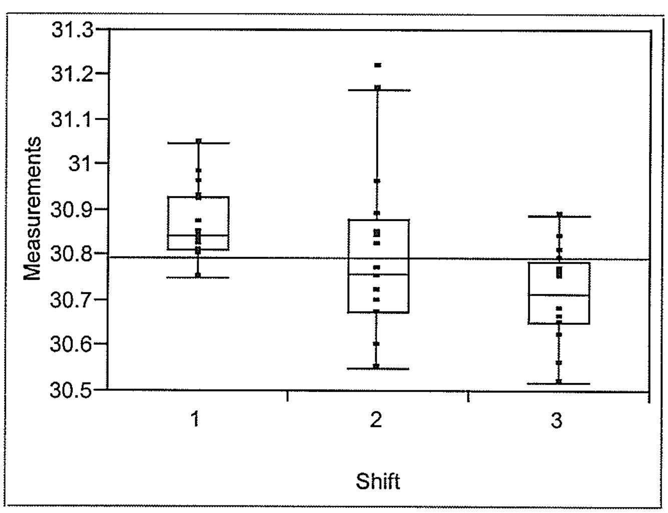


Figure 8.3 Box and whisker plot – Day 7 by Shift

In Figure 8.2 and Figure 8.3, we can see that Shift 2 had a larger variation than the other two shifts on both Day 3 and Day 7. In other words, Shift 2 may be finding it difficult to make consistent measurements. This may well be the reason why Day 3 and Day 7 have larger variations. We will now examine how Shift 2 made measurements over the 4 sites on Day 3 and Day 7. The box and whisker plots for Shift 2 across the 4 sites on Day 3 and Day 7 are displayed in Figure 8.4 and Figure 8.5.

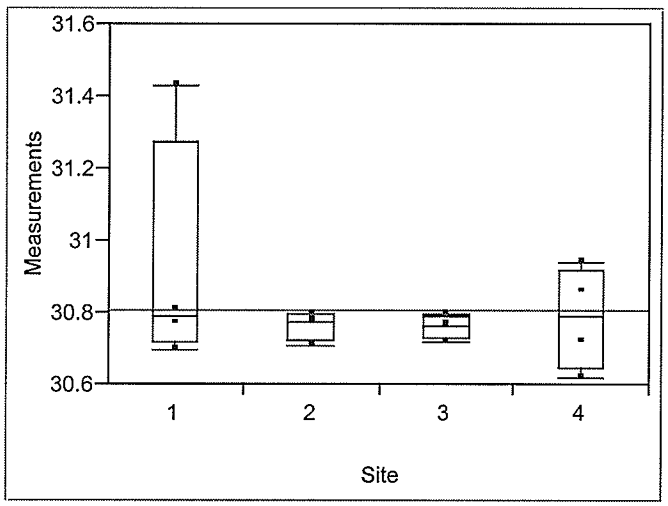


Figure 8.4 Box and whisker plot – Day 3 (Shift 2 by Site)

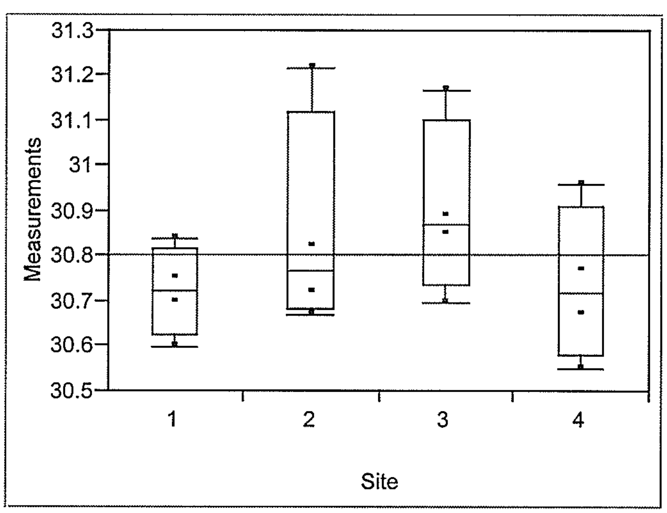


Figure 8.5 Box and whiskers plot – Day 7 (Shift 2 by Site)

In Figure 8.4, we can see that Shift 2 had difficulty making consistent measurements at Site 1. However, we don't see this happening on Day 7. In Figure 8.5, the variations for Site 2, Site 3 and Site 4 are quite large as compared to Site 1. This suggested that Shift 2 had difficulty with consistency of measurements on Day 7, especially at Site 2, Site 3 and Site 4.

Next, we will use the analysis of variance to analyze the gauge capability study with a nested and factorial design. In this example, three factors of interest are Day (D), Shift (O), and Site (S). We treated all three factors as random and the model for this design, as given by Borror, Montgomery and Runger (1997), is

$$X_{ijkm} = \mu + D_i + O(D)_{j(i)} + S_k + SO(D)_{kj(i)} + R_{(ijk)m} \quad \begin{cases} i = 1,2,\cdots,d \\ j = 1,2,\cdots,o \\ k = 1,2,\cdots,s \\ m = 1,2,\cdots,n \end{cases} \tag{8.1}$$

where X_{ijkm} is the m^{th} measurement at the k^{th} site made by the j^{th} shift on the i^{th} day.

Now, using the data from Table II in the Appendix, we can construct the analysis of variance for the model in Equation (8.1). The sums of squares are computed as follows:

$$\begin{aligned} SS_T &= \sum_{i=1}^{d}\sum_{j=1}^{o}\sum_{k=1}^{s}\sum_{m=1}^{n} X_{ijkm}^2 - \frac{X_{....}^2}{dosn} \\ &= (30.69)^2 + (30.70)^2 + \cdots + (30.79)^2 - \frac{(10329.21)^2}{336} \\ &= 2.999 \end{aligned}$$

$$SS_{Day} = \frac{1}{osn}\sum_{i=1}^{d} X_{i\cdots}^{2} - \frac{X_{\cdots\cdot}^{2}}{dosn}$$
$$= \frac{1}{48}[(1475.29)^2 + \cdots\cdots + (1478.22)^2] - \frac{(10329.21)^2}{336}$$
$$= 0.267$$

$$SS_{Site} = \frac{1}{don}\sum_{k=1}^{s} X_{\cdot\cdot k\cdot}^{2} - \frac{X_{\cdots\cdot}^{2}}{dosn}$$
$$= \frac{1}{84}[(2582.68)^2 + \cdots\cdots + (2582.11)^2] - \frac{(10329.21)^2}{336}$$
$$= 0.005$$

$$SS_{O(D)} = \frac{1}{sn}\sum_{i=1}^{d}\sum_{j=1}^{o} X_{ij\cdot\cdot}^{2} - \frac{1}{osn}\sum_{i=1}^{d} X_{i\cdots}^{2}$$
$$= \frac{1}{16}[(491.79)^2 + \cdots\cdots + (491.43)^2] - 317537.705$$
$$= 0.416$$

$$SS_{R} = \sum_{i=1}^{d}\sum_{j=1}^{o}\sum_{k=1}^{s}\sum_{m=1}^{n} X_{ijkm}^{2} - \frac{1}{n}\sum_{i=1}^{d}\sum_{j=1}^{o}\sum_{k=1}^{s} X_{ijk\cdot}^{2}$$
$$= 317540.437 - \frac{1}{4}[(122.86)^2 + \cdots\cdots + (123.14)^2]$$
$$= 1.685$$

In above calculation, SS_R can be considered as $SS_{repeatability}$. This is because, SS_R can be represented as a sum of squares of the differences of observations **within** repeated measurements from their average. The repeatability in this study can be considered as the varability **within** 4 replicated measurements, at 4 specific sites, across 3 shifts, on 7 days.

Finally,

$$
\begin{aligned}
SS_{SO(D)} &= SS_T - SS_R - SS_{Day} - SS_{Site} - SS_{O(D)} \\
&= 2.999 - 1.685 - 0.267 - 0.005 - 0.416 \\
&= 0.626
\end{aligned}
$$

In this example, the expected mean squares for the components are given by

$$
\begin{aligned}
E(MS_{Day}) &= \sigma_R^2 + 4\sigma_{SO(D)}^2 + 16\sigma_{O(D)}^2 + 48\sigma_D^2 \\
E(MS_{Site}) &= \sigma_R^2 + 4\sigma_{SO(D)}^2 + 84\sigma_S^2 \\
E(MS_{O(D)}) &= \sigma_R^2 + 4\sigma_{SO(D)}^2 + 16\sigma_{O(D)}^2 \\
E(MS_{SO(D)}) &= \sigma_R^2 + 4\sigma_{SO(D)}^2
\end{aligned}
$$

and

$$E(MS_R) = \sigma_R^2$$

The analysis of variance is summarized as follows:

Source	DF	Sum of Squares	Mean Square	F-value	P-value
Day	6	0.267	0.0445	1.498	0.2496
Shift (Day)	14	0.416	0.0297	2.856	0.0025
Site	3	0.005	0.0017	0.163	0.9210
Site*Shift (Day)	60	0.626	0.0104	1.552	0.0100
Repeatability	252	1.685	0.0067		
Total	335	2.999			

Table 8.2

Based on the analysis of variance from Table 8.2, we see that the factors Day and Site with P-value 0.2496 and 0.9210 respectively are not significant. However, the Shift within Day effect and the Site crossed with Shift within Day effect do differ significantly.

Thus, the major source of variability is from the shift-to-shift variation within days. Were we to want to reduce this shift-to-shift variability we need to find out why the measurements that are made by different operators show such inconsistency on different days.

We have now to decide which terms should keep in the final model. In Table 8.2, the Day effect and the Site effect are not statistically significant. However, we note that the Site crossed with Shift within Day effect is statistically significant. Because of this, we should keep the Site and the Day effect in our final model. For these reasons, we will retain all four terms (Day, Site, Shift within Day, and Site crossed with Shift within Day) in the final model.

We now show how to use the mean squares from Table 8.2 to obtain the estimates of the variance components. In order to construct estimators for the variance components, we again consider the expected mean squares provided earlier. We can easily see that, the variance component σ_D^2 can be written as

$$\sigma_D^2 = \frac{E(MS_{Day}) - E(MS_{O(D)})}{48}$$

and hence its estimator is

$$\hat{\sigma}_D^2 = \frac{MS_{Day} - MS_{O(D)}}{48}$$

We can also find the estimators for the variance components σ_S^2, $\sigma_{O(D)}^2$, $\sigma_{SO(D)}^2$ and σ_R^2 similarly:

$$\hat{\sigma}_S^2 = \frac{MS_{Site} - MS_{SO(D)}}{84}$$

$$\hat{\sigma}^2_{O(D)} = \frac{MS_{O(D)} - MS_{SO(D)}}{16}$$

$$\hat{\sigma}^2_{SO(D)} = \frac{MS_{SO(D)} - MS_R}{4}$$

and

$$\hat{\sigma}^2_R = MS_R$$

The estimate of the variance component for Day is

$$\hat{\sigma}^2_D = \frac{MS_{Day} - MS_{O(D)}}{48} = \frac{0.0445 - 0.0297}{48} = 0.0003$$

The estimate of the variance component for Site is

$$\hat{\sigma}^2_S = \frac{MS_{Site} - MS_{SO(D)}}{84} = \frac{0.0017 - 0.0104}{84} = -0.0001 \text{ (estimate} < 0\text{)}$$

The estimate of the variance component for Shift within Day is

$$\hat{\sigma}^2_{O(D)} = \frac{MS_{O(D)} - MS_{SO(D)}}{16} = \frac{0.0297 - 0.0104}{16} = 0.0012$$

The estimate of the variance component for Site crossed with Shift within Day is

$$\hat{\sigma}^2_{SO(D)} = \frac{MS_{SO(D)} - MS_R}{4} = \frac{0.0104 - 0.0067}{4} = 0.0009$$

The repeatability of the gauge can be estimated by MS_R, this can be computed as

$$\hat{\sigma}^2_{repeatability} = MS_R = 0.0067$$

Borror, Montgomery and Runger (1997) has shown that the estimates of the variance components for reproducibility and gauge variability can be computed as follows:

$$\begin{aligned}\hat{\sigma}^2_{reproducibility} &= \hat{\sigma}^2_{O(D)} + \hat{\sigma}^2_{SO(D)} \\ &= 0.0012 + 0.0009 \\ &= 0.0021\end{aligned}$$

and

$$\begin{aligned}\hat{\sigma}^2_{gauge} &= \hat{\sigma}^2_{repeatability} + \hat{\sigma}^2_{reproducibility} \\ &= 0.0067 + 0.0021 \\ &= 0.0088\end{aligned}$$

Using the results above, the confidence intervals on $\sigma^2_{repeatability}$ can be easily obtained. The 95% confidence interval on $\sigma^2_{repeatability}$ can be computed as

$$\begin{gathered}\frac{dos(n-1)\hat{\sigma}^2_{repeatability}}{\chi^2_{0.025,dos(n-1)}} \le \sigma^2_{repeatability} \le \frac{dos(n-1)\hat{\sigma}^2_{repeatability}}{\chi^2_{0.975,dos(n-1)}} \\ \frac{252(0.0067)}{297.864} \le \sigma^2_{repeatability} \le \frac{252(0.0067)}{209.923} \\ 0.0057 \le \sigma^2_{repeatability} \le 0.0080\end{gathered}$$

In Chapter 5, we mentioned that the exact confidence intervals for $\sigma^2_{reproducibility}$ and σ^2_{gauge} may not be found. However, Searle (1971) provides details for the construction of approximate $100(1-\alpha)\%$ confidence interval estimates for any linear function of expected mean squares. Using Equation (5.3), an approximate 95% confidence interval on $\sigma^2_{reproducibility}$ can be determined from

$$\frac{v\hat{\sigma}^2_{reproducibility}}{\chi^2_{0.025,v}} \leq \sigma^2_{reproducibility} \leq \frac{v\hat{\sigma}^2_{reproducibility}}{\chi^2_{0.975,v}}$$

where, similar to Equation (5.4) but for the model used in this example,

$$v = \frac{(\hat{\sigma}^2_{reproducibility})^2}{\frac{(1/sn)^2 MS^2_{O(D)}}{df_{O(D)}} + \frac{(s-1/sn)^2 MS^2_{SO(D)}}{df_{SO(D)}} + \frac{(1/n)^2 MS^2_R}{df_R}}$$

$$= \frac{(0.0021)^2}{\frac{(1/16)^2(0.0297)^2}{14} + \frac{(3/16)^2(0.0104)^2}{60} + \frac{(1/4)^2(0.0067)^2}{252}}$$

$$= 13.7543 \approx 14$$

Therefore, the approximate 95% confidence interval on $\sigma^2_{reproducibility}$ is

$$\frac{14(0.0021)}{26.1189} \leq \sigma^2_{reproducibility} \leq \frac{14(0.0021)}{5.62873}$$

$$0.0011 \leq \sigma^2_{reproducibility} \leq 0.0052$$

Similarly, using Equation (5.5) an approximate 95% confidence interval on σ^2_{gauge} can be found as

$$\frac{u\hat{\sigma}^2_{gauge}}{\chi^2_{0.025,u}} \leq \sigma^2_{gauge} \leq \frac{u\hat{\sigma}^2_{gauge}}{\chi^2_{0.975,u}}$$

where u is now given by

$$u = \frac{(\hat{\sigma}^2_{gauge})^2}{\frac{(1/sn)^2 MS^2_{O(D)}}{df_{O(D)}} + \frac{[(s-1)/sn]^2 MS^2_{SO(D)}}{df_{SO(D)}} + \frac{[(n-1)/n]^2 MS^2_R}{df_R}}$$

$$= \frac{(0.0088)^2}{\frac{(1/16)^2(0.0297)^2}{14} + \frac{(3/16)^2(0.0104)^2}{60} + \frac{(3/4)^2(0.0067)^2}{252}}$$

$$= 189.0188 \approx 189$$

Therefore, the approximate 95% confidence interval on σ^2_{gauge} is

$$\frac{189(0.0088)}{228.9636} \leq \sigma^2_{gauge} \leq \frac{189(0.0088)}{152.8221}$$

$$0.0073 \leq \sigma^2_{gauge} \leq 0.0109$$

Appendix

I. Table I: Factors for Constructing Variable Control Charts

II. Table II: Semiconductor Data

Table I: Factors for Constructing Variables Control Chart

Observations in Sample	A	c_4	B_5	B_6	d_2
2	2.121	0.7979	0	2.606	1.128
3	1.732	0.8862	0	2.276	1.693
4	1.500	0.9213	0	2.088	2.059
5	1.342	0.9400	0	1.964	2.326
6	1.225	0.9515	0.029	1.874	2.534
7	1.134	0.9594	0.113	1.806	2.704
8	1.061	0.9650	0.179	1.751	2.847
9	1.000	0.9693	0.232	1.707	2.970
10	0.949	0.9727	0.276	1.669	3.078
11	0.905	0.9754	0.313	1.637	3.173
12	0.866	0.9776	0.346	1.610	3.258
13	0.832	0.9794	0.374	1.585	3.336
14	0.802	0.9810	0.399	1.563	3.407
15	0.775	0.9823	0.421	1.544	3.472
16	0.750	0.9835	0.440	1.526	3.532
17	0.728	0.9845	0.458	1.511	3.588
18	0.707	0.9854	0.475	1.496	3.640
19	0.688	0.9862	0.490	1.483	3.689
20	0.671	0.9869	0.504	1.470	3.735
21	0.655	0.9876	0.516	1.459	3.778
22	0.640	0.9882	0.528	1.448	3.819
23	0.626	0.9887	0.539	1.438	3.858
24	0.612	0.9892	0.549	1.429	3.895
25	0.600	0.9896	0.559	1.420	3.931

Source: Montgomery, D.C. (1997). *Introduction to Statistical Quality Control*, 3rd ed. John Wiley & Sons, New York, NY.

Table II: Semiconductor Data

Day	Shift	Site	Measurements			
1	1	1	30.69	30.70	30.74	30.73
1	1	2	30.73	30.76	30.73	30.76
1	1	3	30.73	30.76	30.76	30.75
1	1	4	30.73	30.74	30.75	30.73
1	2	1	30.68	30.83	30.76	30.96
1	2	2	30.70	30.76	30.71	30.74
1	2	3	30.71	30.72	30.72	30.71
1	2	4	30.71	30.70	30.72	30.72
1	3	1	30.77	30.73	30.72	30.73
1	3	2	30.76	30.71	30.72	30.73
1	3	3	30.76	30.72	30.70	30.71
1	3	4	30.74	30.71	30.72	30.72
2	1	1	30.70	30.70	30.71	30.74
2	1	2	30.68	30.71	30.71	30.71
2	1	3	30.69	30.70	30.71	30.72
2	1	4	30.71	30.68	30.70	30.72
2	2	1	30.72	30.69	30.74	30.75
2	2	2	30.77	30.74	30.73	30.73
2	2	3	30.76	30.74	30.75	30.72
2	2	4	30.75	30.76	30.73	30.72
2	3	1	30.72	30.65	31.02	30.78
2	3	2	30.71	30.67	30.74	30.70
2	3	3	30.70	30.71	30.70	30.72
2	3	4	30.69	30.67	30.69	30.72
3	1	1	31.19	30.60	30.91	30.70
3	1	2	30.71	30.72	30.77	30.73
3	1	3	30.75	30.71	30.77	30.74
3	1	4	30.79	30.76	30.65	30.72
3	2	1	30.81	30.77	30.70	31.43
3	2	2	30.80	30.71	30.78	30.77
3	2	3	30.76	30.80	30.77	30.72
3	2	4	30.86	30.72	30.62	30.94
3	3	1	30.57	30.36	30.58	30.59
3	3	2	30.65	30.67	30.73	30.75
3	3	3	30.76	30.75	30.74	30.76
3	3	4	30.79	30.70	30.66	30.72
4	1	1	30.74	30.89	30.69	30.76
4	1	2	30.73	30.73	30.66	30.68
4	1	3	30.73	30.75	30.70	30.71
4	1	4	30.74	30.70	30.68	30.78
4	2	1	30.64	30.54	30.62	30.77
4	2	2	30.69	30.71	30.66	30.63
4	2	3	30.68	30.75	30.67	30.71
4	2	4	30.76	30.77	30.74	30.70

Table II (Continued)

4	3	1	30.52	30.67	30.67	30.74
4	3	2	30.61	30.69	30.69	30.70
4	3	3	30.70	30.70	30.77	30.60
4	3	4	30.71	30.76	30.76	30.65
5	1	1	30.84	30.71	30.74	30.85
5	1	2	30.71	30.73	30.73	30.76
5	1	3	30.73	30.72	30.75	30.75
5	1	4	30.71	30.76	30.74	30.72
5	2	1	30.93	30.76	30.93	30.81
5	2	2	30.78	30.74	30.79	30.77
5	2	3	30.75	30.73	30.76	30.70
5	2	4	30.76	30.73	30.79	30.76
5	3	1	30.80	30.67	30.65	30.77
5	3	2	30.72	30.83	30.78	30.76
5	3	3	30.72	30.78	30.76	30.82
5	3	4	30.71	30.78	30.75	30.69
6	1	1	30.78	30.53	30.71	30.70
6	1	2	30.75	30.79	30.76	30.80
6	1	3	30.76	30.74	30.73	30.79
6	1	4	30.79	30.81	30.77	30.80
6	2	1	30.71	30.74	30.74	30.61
6	2	2	30.73	30.83	30.71	30.76
6	2	3	30.69	30.58	30.62	30.64
6	2	4	30.73	30.73	30.69	30.61
6	3	1	30.72	30.74	30.72	30.73
6	3	2	30.74	30.75	30.76	30.73
6	3	3	30.74	30.74	30.75	30.75
6	3	4	30.74	30.75	30.72	30.77
7	1	1	30.93	30.82	30.98	31.05
7	1	2	30.82	30.87	30.87	30.96
7	1	3	30.92	30.83	30.81	30.75
7	1	4	30.85	30.84	30.81	30.80
7	2	1	30.84	30.70	30.60	30.75
7	2	2	30.82	30.72	30.67	31.22
7	2	3	30.89	30.70	30.85	31.17
7	2	4	30.77	30.96	30.55	30.67
7	3	1	30.52	30.68	30.66	30.84
7	3	2	30.68	30.68	30.76	30.89
7	3	3	30.56	30.65	30.62	30.75
7	3	4	30.77	30.77	30.81	30.79

Source: Borror, C. M., Montgomery, D.C. and Runger G.C. (1997). "Confidence Intervals for Variance Components from Gauge Capability Studies." *Quality and Reliability Engineering International* 13, pp. 361-369.

Bibliography

1. American Society for Quality Control (1990). *Measurement Systems Analysis Reference Manual.* ASQC, Milwaukee, Wisconsin.
2. Borror, C. M., Montgomery, D. C. and Runger G. C. (1997). "Confidence Intervals for Variance Components from Gauge Capability Studies." *Quality and Reliability Engineering International* 13, pp. 361-369.
3. Deming, W. E. (1982). *Quality, Productivity, and Competitive Position.* Massachusetts Institute of Technology, Center for Advanced Engineering Study, Cambridge, MA.
4. Farnum, N. R. (1994). *Statistical Quality Control and Improvement.* Duxbury Press, Belmont, California.
5. ISO Draft International Standard (2001). *ISO/DIS 10012 – Measurement Control Systems.*
6. ISO Standards Handbook (2000). *Statistical Methods for Quality Control, Volume I, Statistical Methods in General, Terminology and Symbols, Acceptance Sampling.* ISO, Switzerland.
7. ISO Standard Handbook (2000). *Statistical Methods for Quality Control, Volume II, Measurement Methods and Results, Interpretation of Statistical Data, Process Control.* ISO, Switzerland.
8. Juran, J. M. (1979). *Quality Control Handbook,* 3rd ed. McGraw-Hill Inc., New York, NY.

9. Kane, V. E. (1989). *Defect Prevention.* Marcel Dekker, Inc., New York, NY.

10. Maloney, C. J. and Rastogi, S. C. (1970). "Significance Tests for Grubbs's Estimators." *Biometrics* 26, pp. 671-676.

11. Milliken, G. A. and Johnson, D. E. (1984). *Analysis of Messy Data, Volume I; Designed Experiments.* Lifetime Learning Publications, Belmont, California.

12. Montgomery, D. C. and Runger, G. C. (1993a). "Gauge Capability Analysis and Designed Experiments. Part I: Basic Methods." *Quality Engineering* 6(1), pp. 115-135.

13. Montgomery, D. C. and Runger, G. C. (1993b). "Gauge Capability Analysis and Designed Experiments. Part II: Experimental Design Models and Variance Component Estimation." *Quality Engineering* 6(2), pp. 289-305.

14. Montgomery D. C. (2001a). *Design and Analysis of Experiments,* 5th ed. John Wiley & Sons, Inc., New York, NY.

15. Montgomery D. C. (2001b). *Introduction to Statistical Quality Control,* 4th ed. John Wiley & Sons, Inc., New York, NY.

16. Morgan, W. A. (1939). "A Test for the Significance of the Difference Between the Two Variances in a Sample from a Normal Bivariate Population." *Biometrika* 31, pp. 13-19.

17. Searle, S. R. (1971). *Linear Models,* John Wiley & Sons, Inc., New York, NY.

18. Sobralske, B. (1989). "Measuring Up to MIL-STD-45662A." *Quality Progress,* September, pp. 51-52.