

The background is a dark gray field featuring a series of vertical bars of varying heights. To the right of these bars, a series of small, light gray arrows point upwards, creating a sense of growth or progress. The overall aesthetic is modern and technical.

BASIC TOOLS FOR QUALITY IMPROVEMENT IN HEALTH CARE INFORMATICS

JEROME NIYIRORA

Copyright © 2020 **Jerome Niyirora**

This work is licensed under a Creative Commons

Attribution-NonCommercial-ShareAlike 4.0 International Public License.



You are free to:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

NonCommercial — You may not use the material for commercial purposes.

ShareAlike — If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

No additional restrictions — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

ISBN : **978-1-64176-081-2**

Edition 0.0

Cover design by Jerome Niyirora

About the Textbook

This textbook introduces students to the essential tools of quality improvement. The emphasis is placed on health care informatics, as reflected in the several examples contained in the text. The book is written to be accessible to any student in the areas of health information management, health care informatics, and health care industrial engineering. Although having some statistical background would be a plus, such knowledge is not a prerequisite to understanding and applying the tools presented here. Several How-To sections are included to demonstrate the hands-on implementation of the discussed concepts using software such as Minitab, Visio, and Excel.

The textbook starts with **Module 1**, which introduces the general concepts of quality and improvement frameworks. **Module 2** discusses tools for identifying opportunities for improvement. **Module 3** reviews tools for collecting and aggregating data. **Module 4** discusses statistical and graphical tools for assessing the variability and capability of the process. **Module 5** appraises graphical and analytical tools for evaluating relationships in the data. **Module 6** presents tools for assessing the causes and risks of quality problems. **Module 7** introduces tools for benchmarking quality from different processes. **Module 8** reviews tools typically utilized to improve quality. **Module 9** considers tools for monitoring and controlling a process. The Appendix presents factors for constructing variable control charts.

This textbook is not meant to be a comprehensive manuscript of all aspects of quality improvement. Instructors and students can supplement the module reading with additional resources such as those referenced in the bibliography section of this text. The reader is also encouraged to consult a complementary textbook by the same author, titled **Applications of Control Charts for Quality Improvement in Health Care**.

About the Author

Jerome Niyirora, Ph.D. is an Assistant Professor in the Health Information Management (HIM) Program, which is part of the College of Health Sciences at SUNY Polytechnic Institute (SUNY Poly). Jerome graduated from SUNYIT (now SUNY Poly) with a dual bachelor's degree in HIM and Health Services Management. Subsequently, he earned a Master of Science in Health Services Administration. Jerome also holds a Master of Science in Industrial Engineering from SUNY University at Buffalo, and a Ph.D. degree in Systems Science from SUNY Binghamton University. Jerome is a Registered Health Information Administrator (RHIA). His research focuses on the integration of system science theory, calculus of variations, and machine learning techniques into the improvement and optimal design of healthcare processes.

*To
Missy
Jonathan and Aaron*

Contents

1	Key Concepts of Quality Improvement	11
1.1	Introduction	12
1.2	What is quality?	12
1.3	Measuring quality	12
1.3.1	Dimensions of product quality	12
1.3.2	Dimensions of service quality	12
1.3.3	Dimensions of health care quality	12
1.3.4	Quality in health care informatics	14
1.4	Causes of poor quality	15
1.4.1	Human error	15
1.4.2	Poor processes	15
1.4.3	Ineffective information system	15
1.5	Quality and variability	16
1.6	Quality vs. process improvement	16
1.7	Models of quality improvement	17
1.7.1	PDCA	17
1.7.2	DMAIC	18
1.7.3	Technical aspects of quality improvement	19
1.8	Managerial aspects of quality improvement	20
1.8.1	Managerial support	20
1.8.2	Quality costs	21
1.9	A brief history of quality improvement	22
1.9.1	Manufacturing	22
1.9.2	Health care	23

1.10	Check your understanding	24
2	Tools for identifying opportunities for improvement	26
2.1	Introduction	27
2.2	Voice of the customer (VOC)	27
2.3	Quality audits	27
2.4	Baseline measures	28
2.5	Overview of Lean Six-Sigma	29
2.5.1	What is Lean?	29
2.5.2	What is Six-Sigma?	31
2.6	Flowcharts	35
2.6.1	A Basic flowchart	35
2.6.2	A swim-lane diagram	35
2.6.3	A workflow diagram	35
2.6.4	Value-stream map	36
2.7	A project charter and a SIPOC diagram	40
2.7.1	A project charter	40
2.7.2	SIPOC diagram	41
2.8	Check your understanding	42
3	Tools for collecting and presenting the data	45
3.1	Introduction	46
3.2	The need for new data	46
3.3	The types of data to collect	46
3.4	Tools for collecting and organizing data	47
3.5	Tools for presenting the data	48
3.5.1	Checksheets	48
3.5.2	Pie Charts	50
3.5.3	Bar charts	52
3.5.4	Pareto charts	53
3.5.5	Defect concentration diagram	57
3.5.6	Radar charts	57
3.6	Check Your Understanding	59
4	Tools for assessing the process variability and capability	62
4.1	Introduction	63

4.2	Statistical background	63
4.2.1	Process sampling	63
4.2.2	Sample statistics	64
4.2.3	Probability distributions	67
4.2.4	Normal distribution	68
4.2.5	Probability plots	70
4.2.6	Point estimators	73
4.3	Process variability	75
4.3.1	Control charts	75
4.3.2	Histograms	75
4.3.3	Stem-and-leaf diagrams	79
4.3.4	Box plots	82
4.4	Process capability	84
4.4.1	The C_p ratio	84
4.4.2	The C_{pu} , C_{pl} , and C_{pk} ratios	84
4.4.3	Implementing PCRs	85
4.5	Check your understanding	91
5	Tools for evaluating relationships in data	96
5.1	Introduction	97
5.2	Visual tools	97
5.2.1	Scatter diagrams	97
5.2.2	Marginal plots	98
5.3	Statistical measures	100
5.3.1	Correlation analysis	101
5.3.2	Regression	102
5.3.3	Goodness-of-fit	103
5.4	Check your understanding	110
6	Tools for root cause and risk analysis	113
6.1	Introduction	114
6.2	A general procedure	114
6.3	Generating ideas	114
6.4	Organizing ideas	114
6.4.1	Affinity grouping	115
6.4.2	Cause-and-effect diagrams	115
6.4.3	5Whys diagram	117
6.4.4	Why-why diagram	118
6.4.5	Is-Is Not matrix	118

6.5	Prioritizing ideas	119
6.5.1	Voting techniques	119
6.5.2	Failure Modes and Effects Analysis (FMEA)	120
6.5.3	Force-field analysis	122
6.6	Check your understanding	124
7	Tools for benchmarking quality	126
7.1	Introduction	127
7.2	Statistical background	127
7.2.1	Confidence interval	127
7.2.2	Hypothesis testing	130
7.2.3	P-value	131
7.2.4	The type of error	132
7.3	Comparing the means of two independent samples	133
7.4	Comparing the means of more than two independent samples	134
7.5	Comparing the means of two dependent samples	134
7.6	Comparing the means of more than two dependent samples	135
7.7	Comparing variances of two independent samples	137
7.8	ANOVA	138
7.8.1	One-way ANOVA	138
7.8.2	Two-way ANOVA	140
7.8.3	Repeatability and Reproducibility (R&R) Analysis	149
7.9	Chi-square test	154
7.9.1	One variable	154
7.9.2	Two variables	155
7.10	League tables, funnel charts, and ANOM	159
7.10.1	A league table	159
7.10.2	Funnel charts	160
7.10.3	ANOM	160
7.11	Check Your Understanding	166
8	Tools for improving the process	170
8.1	Introduction	171
8.2	DOE	171
8.2.1	Background	171
8.2.2	The procedure for DOE	171
8.2.3	Factorial design	172
8.2.4	Other considerations of factorial designs	176

8.3	Lean tools	183
8.3.1	The cycle time problem	183
8.3.2	Potential solutions	183
8.4	Data quality tools	185
8.5	Prevention tools	186
8.5.1	Poka-Yoke measures	186
8.5.2	Swiss-cheese model	186
8.6	Control charts	186
8.7	Check your understanding	188
9	Tools for monitoring and controlling the process	192
9.1	Introduction	193
9.2	Control charts	193
9.2.1	The Shewhart model	193
9.2.2	How do Shewhart control charts work?	194
9.2.3	Implementing Shewhart control charts	194
9.2.4	How to improve a process using Shewhart control charts?	197
9.2.5	The types of Shewhart's control charts	199
9.2.6	<i>ImR</i> charts	199
9.2.7	A <i>p</i> chart	203
9.3	Run charts	209
9.3.1	A general concept	209
9.3.2	Interpreting run charts	209
9.4	Out-of-control action plan (OCAP)	216
9.5	Dashboards	217
9.6	Transitioning the process back to the owner	219
9.7	Check Your Understanding	220
10	Appendix	227
.1	Appendix	228
	Literature	229

MODULE 1

Key Concepts of Quality Improvement

In this module, we introduce the general concepts of quality improvement. We start by defining the term "quality" and then examine conventional approaches for assessing and improving the quality of products and services. We place particular emphasis on the notions of quality improvement in health care informatics.

Key concepts and tools: *Quality; Variability; Health care; Health care informatics; Donabedian's model; Quality improvement; PDCA cycle; DMAIC framework; Lean-six sigma; Statistical process control; Quality costs; Managerial support*

Major objectives

After studying this module, you will be able to:

- 1. Define quality*
- 2. Describe the general dimensions of quality*
- 3. Determine the dimensions of data quality*
- 4. Explain Donabedian's model of quality*
- 5. Identify the source of quality problems*
- 6. Discuss quality improvement methods*
- 7. Contrast quality and process improvement*
- 8. Appraise quality costs*
- 9. Advocate for managerial support of quality improvement initiatives*
- 10. Explain the history of quality improvement*

1.1 Introduction

We dedicate this module to the discussion of the word “quality”, and attempt to explain what we mean when say “quality improvement.” Here, we present general concepts. In later modules, we will review specific aspects of quality and will learn to utilize standard tools of quality improvement.

1.2 What is quality?

The word “quality” is a broad term that carries different meanings depending on the context. We can think of quality as **the degree of meeting a particular dimension or standard, with little variability**. We establish the relationship between quality and variability in Definition 1.1.

Definition 1.1 (Quality)

Quality inversely relates to **variability** around the standard [43].

1.3 Measuring quality

Following Definition 1.1, we can measure the quality of a product or service on the scale of meeting the set standards or well-defined dimensions. In general, the more variable a product or service, the less likely it will be of good quality.

1.3.1 Dimensions of product quality

A “quality product” is expected to be highly rated in all of the dimensions highlighted in BOX 1.1 [23, 43]:

1.3.2 Dimensions of service quality

The dimensions of “service quality” are not as well defined as those of a quality product. Box 1.2 delineates general features that we use to judge quality in typical services (e.g., retail, banking, etc.).

1.3.3 Dimensions of health care quality

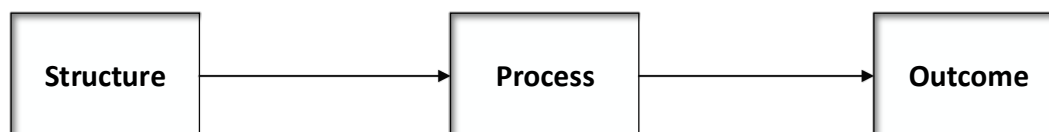
For health care services, we can apply the general dimensions of service quality in BOX 1.2 to evaluate administrative services such as patient registration, billing, bed management, and others. But, we cannot easily apply these dimensions to measure the quality of medical care. For example, we would have difficulties measuring the **reliability** dimension, since, in most cases, we (patients) lack the technical knowledge to determine

Box 1.1 Dimensions of product quality

1. **Performance:** Will the product do the intended job?
2. **Reliability:** How often will the product fail?
3. **Durability:** How long will the product last?
4. **Serviceability:** How easily is the product fixed?
5. **Aesthetics:** What is the visual appeal of the product?
6. **Features:** What more can the product do?
7. **Perceived quality:** What is the reputation of the product or the company?
8. **Conformance to standards:** Is the product made as the designer intended?

the accuracy of care that we receive. This phenomenon is commonly known as **information asymmetry**.¹ Even for experts, the task of measuring the quality of health care is a daunting one given several factors to consider, including social, legal, psychological, behavioral, economic, and technological effects on the quality of care [6]. One of the commonly accepted methods for assessing health care quality is the Donabedian's **Structure-Process-Outcome** model depicted in Figure 1.1.

Figure 1.1: Donabedian's Structure-Process-Outcome model for measuring health care quality



At the **structure** level of Donabedian's paradigm, we measure the efficacy of resources and administrative processes used to provide medical care. At the **process** level, we evaluate the degree to which the applied medical standards are state-of-the-art. At the **outcome** level, we measure the result of care, such as the patient's health status and satisfaction [20]. Specific quality measures that relate to different levels of the Donabedian's model can be found at the websites of the Joint Commission [18] and the Centers for Medicare and Medicaid Services (CMS) [17].

¹**Information asymmetry in health care** is the state of knowledge where patients know less about the medical care they buy than the providers of it [22, 10, 11, 55]

Box 1.2 Dimensions of service quality [52]

1. **Tangibles:** How is the physical appearance of the structure used to deliver a service such as equipment, building, personnel, and communication materials?
2. **Reliability:** Is the service as advertised, and is it accurate?
3. **Responsiveness:** Is the service responsive to the immediate needs of the customers?
4. **Assurance:** Is the service courteous, and do employees convey a sense of trust and confidence?
5. **Empathy:** Does the service exhibit care about the individual needs of the customer?

1.3.4 Quality in health care informatics

The primary functions of a health care informatician involve the capture, storage, retrieval, and analysis of biomedical **data** [5]. For the reader not familiar with health care informatics, the following definition applies for general purposes.

Definition 1.2 (Health care informatics)

Health care informatics is an *“interdisciplinary field that studies and pursues the effective uses of biomedical data, information, and knowledge for scientific inquiry, problem-solving, and decision making, motivated by efforts to improve human health [5].”*

On the Donabedian’s model, we can classify health care informatics services at the *structure* level. To assess the quality of these services, we apply the dimensions in Box 1.2. To evaluate the reliability feature, we, for instance, can measure the degree to which biomedical data exhibit the following general characteristics of data quality [28, 59]:

1. **Uniqueness:** data are stored in a database where each entity only exists once.
2. **Accuracy:** the captured data represent a real-life object according to the data source authority.
3. **Consistency:** data about the same objects are consistent across entities. Inheritance rules of super-to-subtypes are also consistently enforced.
4. **Completeness:** all the required fields are not null.
5. **Timeliness:** the time from capture to availability for use is minimal. The valid period of data is given.

6. **Currency:** data reflect the current standards of the real-life object.
7. **Conformance:** data are used according to the specifications in the metadata.
8. **Referential integrity:** primary to foreign keys constraint rules are enforced, including the rules of insertion, deletion, and updates.

1.4 Causes of poor quality

1.4.1 Human error

“To err is human”[21]. Some of the leading causes of human errors that can translate into poor service quality include:

1. **Planning:** poor planning leads to poor outcomes [55].
2. **Storage:** due to a multitude of reasons, a human operator may forget how to carry out the original plan. Some of the contributing factors to this problem include interruptions and multi-tasking [55].
3. **Execution:** in this instance, a human operator is unable to carry out a well-planned strategy due to the lack of knowledge or other factors [29].

1.4.2 Poor processes

While it is true that, as humans, we are prone to making mistakes, poor processes are generally to blame. In service, characteristics of poor processes include too many hand-offs, unnecessary waiting, and setup problems [24]. Deming, one of the pioneers of quality improvement, has concluded that poor processes are entirely the responsibility of management, not that of the operator [43]. In Deming’s view, without the commitment of the executives, poor processes are likely to persist. We will come back to Deming’s viewpoints later.

1.4.3 Ineffective information system

An ineffective information system can lead to inadequate quality of products and services. Signs of an ineffective information system include [33, 28]:

1. Poor user interfaces
2. Interoperability problems among computer systems
3. Poor data capture controls
4. Missing or outdated data and information governance policies
5. Legacy systems with accessibility issues

6. Redundant data storage and inconsistent metadata
7. Lack of controls over external data sources
8. Lack of management commitment to data quality

In health care, an ineffective information system could not only lead to data quality issues but also poor medical care [61]. In unfortunate situations, patient harm can result [33].

1.5 Quality and variability

The implication of Definition 1.1 is that too much variability is associated with **low quality**, whereas little variability relates to **high quality**.

Definition 1.3 (Quality improvement)

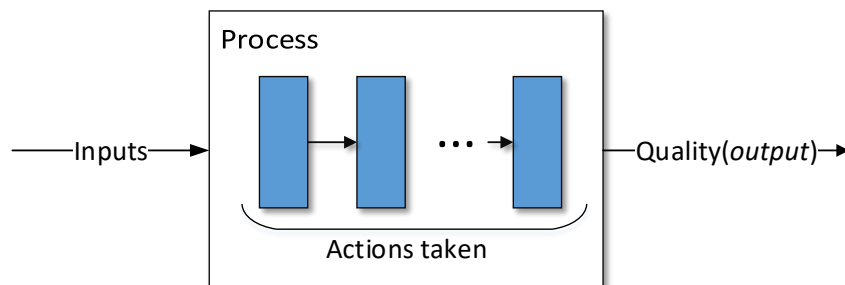
Quality improvement is the **reduction of variability** around the desired standard [43].

The practical significance of Definition 1.3 is that we must understand what the **customer** expects, or is willing to pay for, and what the process variability is before we can improve the quality of a product or service. In medical care, Definition 1.3 requires us to stay informed about the latest evidence-based standards to improve outcomes of care. The implication in health care informatics is that we must practice modern standards of data quality to ensure reliable services.

1.6 Quality vs. process improvement

In this book, we use expressions “quality improvement” and “process improvement” interchangeably to mean the same thing. The justification is that to improve quality, we must improve the related process. Graphically, the relationship between “quality” and “process” is portrayed in Figure 1.2.

Figure 1.2: A model of the relationship between **process** and **quality**



From Figure 1.2, we deduce that “quality,” which is a function of the degree to which the output meets the standard, is dependent on the actions that are taken in a “process.” To **improve quality**, we must **improve the process** by taking a different set of actions that lead to little variability in the outputs. In practice, we design experiments and apply statistical methods to find the best set of actions that minimizes variability. The inputs are not generally the source of variability, especially in manufacturing, where raw materials tend to be standardized. In health care, some inputs, such as risk factors of patients, may affect outcomes of care, thus necessitating adjustments in our quality measures [4].

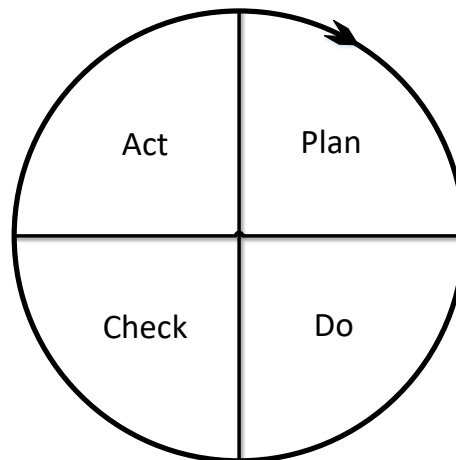
1.7 Models of quality improvement

The most commonly used models of quality improvement are the **Plan, Do, Check, Act (PDCA)** cycle, and the **Define, Measure, Analyze, Improve, Control (DMAIC)** framework.

1.7.1 PDCA

Figure 1.3 portrays a graphical representation of the PDCA cycle, also called the **She-whart cycle** or **PDSA** (where S stands for Study).

Figure 1.3: The PDCA cycle



The arrow on the circle signifies the concept of “continuous” quality improvement. In the **Plan** phase of this cycle, we propose changes or experiments to run. In the **Do** phase, we carry out our plan, usually on a small scale. In the **Check** phase, we analyze the results from the Do phase. In the **Act** phase, we fully implement changes or abandon them. The cycle is iterated as many times as necessary to improve quality [43].

1.7.2 DMAIC

Figure 1.4 shows the basic phases of the DMAIC framework.

Figure 1.4: Phases of the DMAIC framework



The definition of each one of these phases follows [14, 24, 43].

Define: In this phase, we define opportunities for improvement. Additionally, we create flow charts, a project charter,² and a SIPOC³ diagram.

Measure: In this phase, we collect data on key process input variables (KPIV) and key process output variables (KPOV). We attempt to establish baseline measures for WIP,⁴ VOC,⁵ PCE,⁶ cycle time, process completion rate, sigma level, and financial metrics. At the end of the project, we will refer back to these initial measures to assess how much improvement we have made.

Analyze: In this phase, we try to make sense of the data we collected in the Measure phase by analyzing root cause, correlation, risk, stability, capability, and hypothesis tests. One of the tools that we may apply here is the cause-and-effect diagram to evaluate the causes of the quality problem. Another essential tool that we may deploy is a preliminary control chart to assess the stability of the process.

Improve: This phase is probably the most difficult since we have to turn our analysis into actions to improve the process. Creative thinking is imperative. To generate ideas for improvement, we may have to hold brainstorming sessions, not only with current process stakeholders but also with experts and the crowd [40]. One particularly useful statistical tool that we can employ here is DOE⁷ [43]. For problems related to efficiency and cycle times, we can employ **Lean** techniques to eliminate non-value-added activities from the process. We continue to apply these improvement tools until control charts signal that the process is stable.

²**Project charter:** a short document with clear statements about the goal, the business case, the opportunity, the project scope, plan, and team for quality improvement [43]

³**SIPOC:** Suppliers, Inputs, Process, Outputs, and Customers

⁴**WIP:** Work-in-Process (e.g., work yet to be finished)

⁵**VOC:** Voice of the Customer (e.g., customer feedback)

⁶**PCE:** Process Cycle Efficiency

⁷**DOE:** Design of Experiments

Control: In this phase, we implement tools to monitor the process, such as control charts, run charts, and dashboards. To the extent possible, we should automate this phase. Otherwise, we must develop a sampling plan to audit and monitor the new process. Additionally, we must create an out-of-control plan (OCAP) to prescribe actions to be taken when out-of-control behaviors arise in the improved process.

1.7.3 Technical aspects of quality improvement

We have previously mentioned that we cannot improve quality if we don't know the standard. If the standard is not given, we need to know how to approximate it from the process samples. Additionally, we must be able to measure variability if we are to reduce it to improve quality. All these activities require technical skills that, without them, quality improvement efforts may fail. Indeed, it is believed that the movement of total quality improvement (TQM) in the 1980s was not overly successful, partly because of the lack of emphasis on statistical aspects of quality improvement [43]. The strategy that tends to emphasize technical skills more is **Lean Six-Sigma**. The Lean part of this strategy focuses on making processes efficient. The **Six-Sigma** part utilizes statistical techniques to reduce variability in the process to reach a performance state of no more than 3.4 defective parts per million (PPM) opportunities (DPMO) in the process [24]. **Statistical process control (SPC)** is another strategy that emphasizes the reduction of variability in the process to achieve stability and capability to meet quality standards. The seven primary tools of SPC, which are also encountered in the Lean Six-Sigma strategy, are [43]:

1. **Histograms or stem-and-leaf plots** to visually assess the distribution of the process given the central tendency and variability. These tools may also help evaluate the process capability to meet the given quality specifications.
2. **Checksheets** to help collect and tabulate the process data by frequency, time, location, and cause.
3. **Pareto charts** to identify the "vital few" factors causing the majority of quality problems. A common analytical method applied here is the 80/20 rule that implies that 20% of the factors cause 80% of the problems.
4. **Cause-and-effect diagrams** to help analyze the root causes of the given effect. The causes are often categorized by factors such as materials, machines, measurement, people, methods, and policy.
5. **Defect concentration diagrams** to model the spatial distribution of defects on a particular product.
6. **Scatter diagrams** to characterize the relationship between two variables in the process. We often seek to conclude positive correlation, negative correlation, or no correlation between the variables under study.

7. **Control charts** to monitor the process behaviors over time or by the sample number. Control charts help detect and reduce special cause variation in the process.

We will discuss SPC tools further in later modules.

1.8 Managerial aspects of quality improvement

1.8.1 Managerial support

Successful quality improvement programs require not only the technical know-how but also the buy-in from executives. Indeed, without a meaningful managerial involvement, many quality improvement efforts are likely to fail. Deming believed that the success of any quality improvement program was a direct consequence of the actions of senior managers [43]. Deming often spoke of the following **seven deadly diseases of management** [43]:

1. Lack of constancy of purpose
2. Emphasis on short-term profits
3. Evaluation of performance, merit rating, and annual reviews of performance
4. Mobility of top management
5. Running a company of visible figure alone
6. Excessive medical costs
7. Excessive legal damage awards.

Deming recommended that senior managers adopt the following **14 points** [43, 44, 25, 45]:

1. **Create constancy of purpose** to improve the quality of products and services.
2. **Adopt a new philosophy** that recognizes the state of a new economic era.
3. **Do not rely on mass inspection** to control quality (e.g., focus on prevention through process improvement).
4. **Do not award business contracts to suppliers by price alone** without factoring in the costs of quality.
5. **Focus on continuous improvement** of products and services.
6. **Practice modern training** methods (e.g., everyone needs to be trained to practice modern quality improvement methods).
7. **Improve leadership and supervision methods** (e.g., supervisors should focus on helping operators improve the quality of products and services).

8. **Drive out fear** (e.g., let everyone speak out about quality issues)
9. **Break down the barriers between functional areas of the business** (e.g., encourage teamwork).
10. **Eliminate targets and slogans** for the workforce (e.g., targets and slogans are meaningless without commitment and a plan to achieve quality excellence).
11. **Eliminate numerical quotas and work standards** (e.g., quotas and standards without regard to quality are useless).
12. **Remove the barriers** that discourage employees from doing their jobs.
13. **Institute an ongoing program of education** for all employees (e.g., have a continuous education program for all employees to learn how to use quality improvement tools, including statistical methods).
14. Create a managerial structure on the top that will **vigorously support the first 13 points**

1.8.2 Quality costs

It is common practice in many organizations that before we can start working on an expensive quality improvement project, we justify the need for the project and make a business case of our plan. Depending on the financial return, senior managers may deny our project. This concept is demonstrated in the next example.

Example 1.1 A proposed quality improvement project is expected to increase the revenue of Metropolis Hospital by 1%, increase medical expenses by 0.5%, and add \$500,000 in administrative costs. What will be the impact on the hospital's net income if its total revenues, total medical expenses, and overall administrative costs are currently \$100,000,000, \$70,000,000, and \$25,000,000, respectively? A simplified Profit & Loss (P&L) statement follows.

Table 1.1: A P&L Statement of Metropolis Hospital, 2018

	Current	Pro-forma
Revenue	\$100,000,000	\$101,000,000
Expenses		
Medical	\$70,000,000	\$70,350,000
Administrative	\$25,000,000	\$25,500,000
Total expenses	\$95,000,000	\$95,850,000
Net income	\$5,000,000	\$5,150,000

How is the top management likely to respond to this request? Given the likely increase of 3% in the net revenue $((5,150,000 - 5,000,000)/5,000,000 = 0.03)$, the proposal is likely to be approved.

An organization committed to quality should consider all aspects of **quality costs** when deciding on the merits of an improvement project. Quality costs are typically classified in the following categories [43]:

1. **Prevention costs:** costs that go into the efforts of averting defects at the design or operation level.
2. **Appraisal costs:** costs associated with auditing and inspecting products and services to ensure conformance to standards.
3. **Internal failure costs:** costs due to discovered defects before products and services are delivered to the customer. These costs may result from scraps and rework.
4. **External failure costs:** these are costs that go into the efforts of investigating and handling substandard products and services that were delivered to the customer. Examples of such costs include returned products, warranty, liability, bad reputation, and legal fees.

1.9 A brief history of quality improvement

1.9.1 Manufacturing

Quality improvement in manufacturing can be traced as far back as at the start of the industrial revolution in the 1700s. From 1700 to 1900, improvement efforts were mostly decentralized and confined to each manufacturer. By the early 1900s, ideas for improving manufactured products became well-accepted, and they evolved into modern quality improvement techniques that we still use today. To name a few examples, between 1922 and 1923, a statistician named Ronald Fisher published the concepts of DOE that we discussed earlier. In 1924, Walter A. Shewhart introduced control charts for process monitoring. By the 1930s, Henry Ford had developed practical approaches to mistake-proofing assembly lines. In the 1940s, W. Edwards Deming was giving lectures on statistical quality control and quality management. In 1950, Kaoru Ishikawa introduced the cause-and-effect diagram. In 1951, Armand Feigenbaum published the first edition of his book on *Total Quality Control*. Feigenbaum is credited for publicizing the importance of managerial leadership in quality improvement. By mid-1950s, Joseph Juran, who had worked for Shewhart, had become a great champion of quality control and was also giving lectures on quality improvement and management. Juran is also known for the *Juran Trilogy* of quality management, which includes: planning, control, and improvement. In 1959, S.W Roberts extended the idea of control charts into exponentially weighted moving average charts. In 1960, Kaoru Ishikawa introduced the ideas of quality control circles to encourage workers who were doing the same job to meet regularly to attempt to solve common problems. By the mid-1970s, the ideas of quality circles had evolved into the concept of total quality improvement (TQM). In the late 1980s, Motorola publicized its notion of Six-Sigma quality using the idea of 3.4 DPMO [43]. By the early 2000s, many

other ideas of quality improvement had flourished, including Lean and Six-Sigma [24]. During this period of the evolution of quality improvement, there were also many government and private sector initiatives that were critical in advancing the theory and practice of quality improvement [43].

1.9.2 Health care

In health care, quality improvement ideas can be traced as far back as the early 1800s with Dr. Ignaz Semmelweis. Dr. Semmelweis advocated for many ideas of quality, including hand washing. By the 1850s, Florence Nightingale was already a champion of quality of care. Nightingale was known to use many statistical concepts such as polar diagrams to establish correlations between living conditions and mortality rates [50]. In the 1890s, John Hopkins introduced hospital quality standards and initiatives for extensive training of physicians. By the early 1900s, Earnest Godman was a strong advocate of patient follow-ups and public report cards on quality [55]. In 1966, Avedis Donabedian published his famous paradigm of *structure-process-outcome* to measure health care quality. During the period that followed, many government agencies and private organizations were created to promote the quality of care. Examples of these organizations include the Institute of Medicine (IOM), Joint Commission, Professional Standards Review Organizations (PSROs), Agency for Healthcare Research and Quality (AHRQ), National Committee for Quality Assurance (NCQA), The National Quality Forum (NQF), Surgical Care Improvement Program (SCIP), and several others [39]. Today, many of the quality improvement techniques that were developed in the manufacturing sector are widely used in health care to improve the quality of care and administrative services [43, 61].

1.10 Check your understanding

1. Who introduced control charts?
 - (a) Ishikawa
 - (b) Shewhart
 - (c) Juran
 - (d) Donabedian
 - (e) Feigenbaum
2. Who introduced the structure-process-outcome paradigm to measure health care quality?
 - (a) Ishikawa
 - (b) Shewhart
 - (c) Juran
 - (d) Donabedian
 - (e) Feigenbaum
3. Where would you classify health care informatics on the structure-process-outcome model?
4. *Data use according to the specifications in the metadata* relates to which dimension of data quality?
 - (a) Accuracy
 - (b) Uniqueness
 - (c) Conformance
 - (d) Currency
5. *Storing data in a database so that each entity exists only once* is an example of which dimension of data quality?
 - (a) Accuracy
 - (b) Uniqueness
 - (c) Conformance
 - (d) Currency
6. You have noticed the *lack of confidence* in your employees. This likely indicates quality problems in which dimension of service quality?
 - (a) Tangibles
 - (b) Reliability

- (c) Responsiveness
 - (d) Assurance
 - (e) Empathy
7. What happens to quality when variation increases?
 8. What is the basic definition of “quality improvement”?
 9. How does *quality improvement* differ from *process improvement*?
 10. What role does a customer play in quality improvement efforts?
 11. What do we do in the Check phase of PDCA?
 12. What are the dimensions of service quality in health care informatics?
 13. What are the prerequisites of a successful quality improvement program in an organization?
 14. How would you classify costs associated with yield loss due to system downtime?
 - (a) Internal failure costs
 - (b) External failure costs
 15. How would you classify quality costs associated with the investigation of customer complaints?
 - (a) Internal failure costs
 - (b) External failure costs
 16. Who were the first pioneers and champions of quality management?
 17. What are the quality ramifications of a poor health information system?
 18. How would you go about assessing the quality of health care informatics services?
 19. What are the common sources of quality issues?
 20. A proposed quality improvement project about data quality at Metropolis Hospital is expected to increase medical expenses by 0.5%, and add \$500,000 in administrative costs. No change in revenue is expected. What will be the impact on the hospital's net income if its total revenues, total medical expenses, and overall administrative costs are currently \$100,000,000, \$70,000,000, and \$25,000,000, respectively? How is the hospital likely to prioritize this project request? Justify your response and discuss the implications of various quality costs.

MODULE 2

Tools for identifying opportunities for improvement

In this module, we explore essential tools for identifying opportunities for improvement in a process. Among the tools we discuss, include VOC¹, flowcharts, Lean Six-Sigma measures, quality audits, a project charter, and a SIPOC² diagram. We employ Microsoft Visio and Excel to implement these tools.

Key concepts and tools: VOC; Flowchart; Project charter; SIPOC; PCE³; Sigma level; DPMO⁴; Efficiency; Variability; Quality audits

Major objectives

After studying this module, you will be able to:

- 1. Define key concepts and tools for identifying a quality problem*
- 2. Explain the importance of VOC*
- 3. Promote the need for quality audits*
- 4. Construct a basic flowchart*
- 5. Construct a swim-lane diagram*
- 6. Create a value-stream map*
- 7. Draw a spaghetti diagram*
- 8. Calculate Lean Six-Sigma baseline measures*
- 9. Develop a project charter*
- 10. Create a SIPOC diagram*

¹**VOC:** Voice of the customer

²**SIPOC:** Suppliers, Input, Process, Output, Customer

³**PCE:** Process cycle efficiency

⁴**DPMO:** Defects per million opportunities

2.1 Introduction

During the *Define* phase of DMAIC⁵, we attempt to identify opportunities for quality improvement. An important tool that we deploy in this phase is VOC, where we attempt to find opportunities for improvement from the feedback of the customer. We also create **flowcharts** and conduct **quality audits** to identify deficient areas in the process. Once we know an opportunity for improvement exists, we create a **project charter** and a **SIPOC**⁶ diagram to initiate a quality improvement project. We also establish baseline measures using various metrics such as the sigma level, DPMO, and PCE.

2.2 Voice of the customer (VOC)

The process of VOC involves obtaining feedback from customers about the quality of service or product that they've received or would desire. The goal is to understand the priorities and standards of the customers. Typical activities of VOC include interviews, satisfaction surveys, and focused groups [43]. In the early stages of the design of a product or service, we can use VOC for DFSS⁷ to create a product or service that will need little to no improvement in the future. In addition to the traditional approaches of VOC, **health care** organizations may also acquire feedback from professional, regulatory, and accrediting agencies. For example, a hospital could use accreditation outcomes or publicly available quality measures [17] to identify potential opportunities for improvement.

Example 2.1 (Pneumonia care)

The Centers for Medicare and Medicaid Services (CMS) recommends that immunocompetent patients with community-acquired pneumonia receive an antibiotics regimen within 24 hours of admission to the hospital [16]. A hospital could use this guideline as VOC to improve its pneumonia care.

In **health care informatics**, the approach of VOC generally entails requesting feedback from internal customers (e.g., medical and administrative users) about data quality and other information system needs.

2.3 Quality audits

Besides VOC surveys, we can also find opportunities for improvement by performing regular audits or inspections of the process. In **health care informatics**, audits are a must to ensure data quality and compliance with standards (e.g., HL7⁸) and regulations (e.g., HIPAA⁹). Activities in such audits may include [28]:

⁵**DMAIC**: Define, Measure, Analyze, Improve, Control

⁶**SIPOC**: Suppliers, Input, Process, Output, Customer

⁷**DFSS**: Designing for Six-Sigma

⁸**HL7**: Health Level Seven [8]

⁹**HIPAA**: Health Insurance Portability and Accountability Act [13]

1. Data profiling to look for unexpected extreme values or unusual patterns.
2. Comparing data against the business rules.
3. Looking for overwritten warnings.
4. Checking for the validity of data.
5. Looking for duplicated data.
6. Inspecting data entry controls.
7. Examining access and error logs.

Additionally, we should scrutinize the reliability of data entries in our EHRs¹⁰ to ensure [59]:

1. **Completeness:** did our EHR capture all “truth” about the patient?
2. **Correctness:** are the data elements in our EHR true about the patient?
3. **Concordance:** do the data in our EHR agree with other data sources?
4. **Plausibility:** do our data elements look reasonably true, given the measure of interest?
5. **Currency:** how relevant is the data given the time and patient’s state?

It is important to note that Deming has cautioned about relying on mass inspection as a way to improve quality [43]. We should always carry out audits as a part of a broader continuous quality improvement program to find and eliminate all sources of variability in the process.

2.4 Baseline measures

Before we can improve a process, we must be able to measure the gap between the current and the desired performance. We calculate baseline measures to gauge the state of the quality in the process and identify opportunities for improvement. The types of baseline measures depend on the process, but sample statistics such as the **mean** and **variance** are standard measures for processes that generate metric data. Other statistics that we can apply to establish baseline measures include the **sigma level**, **DPMO**, and **PCE**. These statistical measures originate from the Lean Six-Sigma quality improvement strategy, as we discuss next.

¹⁰**EHR:** Electronic Health Record

2.5 Overview of Lean Six-Sigma

2.5.1 What is Lean?

The **Lean** part of the Lean Six-Sigma strategy emphasizes **efficiency** by attempting to eliminate **waste** (or **muda**¹¹) out of the process. More generally, muda is any process step that the customer would not be willing to pay for. We also refer to this step as non-value-added (NVA). In service, examples of NVA activities include [24]:

- **Overproduction:** the production of service beyond what is needed for immediate use (e.g., excessive reports and paperwork). This activity will likely increase **cycle time**¹².
- **Unnecessary waiting:** any extra time between the end of one step and the start of the next step is NVA (e.g., waiting for signatures or approvals).
- **Unnecessary transportation:** inefficient transportation methods increase cycle time. Examples include people always walking around to collect or deliver materials.
- **Needless motion:** unnecessary motions increase cycle time and the chances for errors. Examples of such incidents include switching between different computer desks and performing too many keystrokes to accomplish a simple task.
- **Too much inventory:** having more than needed inventory increases cycle time, such as accepting more orders than one can process in time.
- **Overprocessing:** trying to add more value to service more than the customer requires (e.g., excessive inspections). Any extra and unnecessary steps in the process are expensive since they delay the main service.
- **Defect:** any aspect of service that does not conform to the requirements or standards of the customer.

We apply **Lean** tools to strive to have a process with minimal to no NVA activities. By eliminating NVA steps, the cycle time will decrease, thus making the process more efficient. We can measure the level of efficiency using the following process cycle efficiency (PCE) formula:

$$PCE = \frac{\text{VA time}}{\text{Cycle time}} \quad (2.1)$$

where VA stands for **value-added** and $0 \leq PCE \leq 1$. The higher the PCE measure, the more efficient the process. Given that

$$\text{cycle time} = \text{VA time} + \text{NVA time}, \quad (2.2)$$

¹¹**Muda:** a Japanese word meaning “waste”. In *Lean* terminology, the word muda is commonly used to indicate non-value-add activities [53]

¹²**Cycle time:** The average amount of time it takes to complete a service unit or product [24]

we can rewrite Equation 2.1 like this:

$$PCE = \frac{VA \text{ time}}{VA \text{ time} + NVA \text{ time}} \quad (2.3)$$

From Little's law [38], the following relationship also applies:

$$\text{Cycle time} = \frac{WIP}{\lambda} \quad (2.4)$$

where WIP stands for the work-in-process or the amount of work yet to be completed [43], and λ is the completion rate (e.g., productivity rate). By combining Equations 2.1 and 2.4, we observe that:

$$PCE = \frac{VA \times \lambda}{WIP} \quad (2.5)$$

We use the \times symbol to indicate multiplication. Since $0 \leq PCE \leq 1$, Equation 2.5 implies that:

$$0 \leq VA \times \lambda \leq WIP \quad (2.6)$$

From Equations 2.1-2.5, it follows that if we held all other variables constant, we could increase PCE using one or a combination of the following techniques:

1. **Increase the share of VA time** by decreasing the share of NVA time (e.g., remove NVA steps from the process).
2. **Increase the completion rate** by adding resources (e.g., more employees or computer servers).
3. **Decrease WIP** by having our process operate on a *Pull* instead of a *Push* basis, meaning that before accepting new work, we must first finish existing work [24].

Example 2.2 (PCE measure)

A medical coding and billing company has, on average, 30 claims in the billing queue and, the average processing time of each claim is 2 hours. Per Equation 2.4, we can obtain the completion rate λ by:

$$\text{Cycle time} = \frac{WIP}{\lambda} = 2 = \frac{30}{\lambda} \implies \lambda = \frac{30}{2} = 15 \text{ claims per hour} \quad (2.7)$$

Assuming that 30 minutes out of the 2 hours processing time is NVA, we can obtain VA time as follows by $2 - 0.5 = 1.5$ hours, and the PCE measure by Equation 2.5, as follows:

$$PCE = \frac{VA \times \lambda}{WIP} = \frac{1.5 \times 15}{30} = 75\% \quad (2.8)$$

Similarly, we can obtain the same measure using Equation 2.1 like this:

$$PCE = \frac{VA \text{ time}}{\text{Cycle time}} = \frac{1.5}{2} = 75\% \quad (2.9)$$

The manager of this process can increase the PCE measure by attempting to eliminate NVA activities from the process, or by increasing the completion rate by increasing resources. The choice of decreasing WIP (e.g., limiting customer requests) makes business sense only if the marginal cost of an additional unit of resource is more than the corresponding marginal revenue rate.

2.5.2 What is Six-Sigma?

The **Six-Sigma** part of the Lean Six-Sigma strategy focuses on eliminating opportunities for defects and reducing variation in products.

Theoretical Six-Sigma level

The word “Six-Sigma” typically signifies 6 standard deviations (6σ) away from the mean (μ), and this is denoted as $\mu \pm 6\sigma$. In normally distributed data, we expect 68.27% of the data to fall within $\mu \pm 1\sigma$, 95.45% to fall within $\mu \pm 2\sigma$, 99.73% to fall within $\mu \pm 3\sigma$, 99.9937% to fall within $\mu \pm 4\sigma$, 99.999943% to fall within $\mu \pm 5\sigma$, and 99.9999998% to fall within $\mu \pm 6\sigma$. Theoretically, a Six-Sigma process should only have 0.002 defective parts per million (PPM) opportunities (DPMO), which translates into 99.9999998% defect-free parts. In other words, the probability of observing a defective part (or a nonconforming step in the process) is $0.002/1,000,000 = 0.0000002\%$.

Approximating DPMO

Let's assume that the probability of a nonconforming step in the process is p . We can calculate p this way:

$$p = \frac{D}{TOP} \quad (2.10)$$

where D is the total number of nonconforming steps, and TOP is the total number of opportunities for nonconformity given by:

$$TOP = U \times OP \quad (2.11)$$

Here, OP is the number of opportunities per step, and U is the total number of the process steps [14]. Using p , we can estimate DPMO this way:

$$DPMO = p \times 1,000,000 \quad (2.12)$$

Yield in the process

The ratio of the total number of defective units U_D and the total number of units U quantifies the percent of defective units. From this ratio, we can estimate the process yield Y as follows:

$$Y = 1 - \frac{U_D}{U} \quad (2.13)$$

Equation 2.13, which ignores opportunities of more than one defect per unit, is a simplistic way of determining the accuracy or success rate in typical processes.

Example 2.3 (Sigma level)

A health care informatics process requires 20 steps, and each step is considered an opportunity for a mistake. If the accuracy rate in each step is 95%, it follows that the probability of delivering a service unit with no mistake, or the yield, is given by:

$$0.95^{20} = 35.85\% \quad (2.14)$$

In other words, Equation 2.14 indicates that about 64.15% (1-35.85%) of the delivered services are likely to have at least one mistake. So, out of 1,000,000 services, 641,500 of them will be nonconforming. If the goal was to have a yield of 95%, then the accuracy level in each $(1 - p)$ must be:

$$(1 - p)^{20} = 0.95 \quad (2.15)$$

$$= 0.95^{1/20} \quad (2.16)$$

$$= 0.997439 \quad (2.17)$$

$$= 99.74\% \quad (2.18)$$

So, the probability of a step in the process being defective is $p = 1 - 99.74\% = 0.26\%$, which is equivalent to a theoretical 3σ performance level. We can already appreciate that a 6σ performance level practically means no errors in any step of the process.

Motorola Six-Sigma level

In practice, we rarely apply the theoretical Six-Sigma concept, but instead, use the idea of Six-Sigma that Motorola introduced in the 1980s. From Motorola's experience, a product could *drift* by as much as $\pm 1.5\sigma$ from the target and yet remained acceptable. After accounting for such drift, Motorola developed a measure of Six-Sigma based on 3.4 DPMO instead of the theoretical 0.002 DPMO discussed earlier [43]. We denote Motorola Six-Sigma metric as $6\sigma_M$ and calculate it as follows [14]:

$$6\sigma_M = 0.8406 + \sqrt{29.37 - 2.221 * \ln(DPMO)} \quad (2.19)$$

Here, $\ln()$ symbolizes the natural logarithm. We can also obtain the $6\sigma_M$ measure using Excel as we demonstrate in How-To 2.1.

How-To 2.1 (Calculating $6\sigma_M$ in Excel 2013)

To calculate $6\sigma_M$ in Excel, use

$$= ABS(NORMSINV(p)) + 1.5\sigma \quad (2.20)$$

where $+1.5\sigma$ is the drift, $ABS()$ is the absolute value function, and $NORMSINV()$ is an Excel function to calculate the inverse of a standard normal.



Example 2.4 ($6\sigma_M$)

We use the formula in Equation 2.20 to verify that 3.4 defects in a million opportunities translate into $6\sigma_M$. First, we obtain p by:

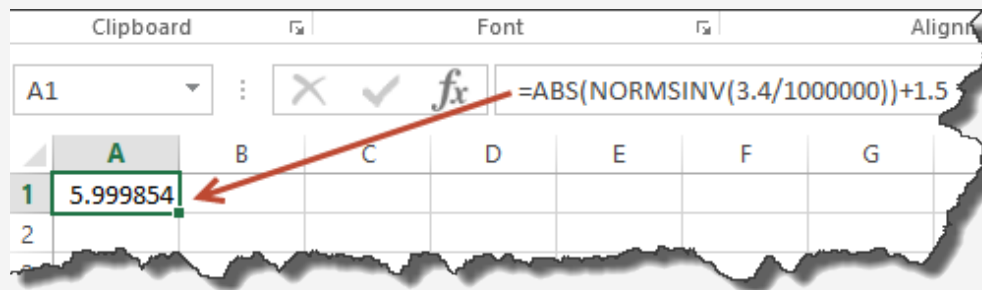
$$p = \frac{3.4}{1,000,000} \quad (2.21)$$

We open Excel and type into cell A1 this formula:

$$=ABS\left(NORMSINV\left(\frac{3.4}{1000000}\right)\right)+1.5 \quad (2.22)$$

which yields $5.999854 \approx 6$.

Figure 2.1: A snapshot of Excel to show the calculation of $6\sigma_M$

**Example 2.5 ($6\sigma_M$ and DPMO)**

As the manager of the revenue cycle at Metropolis Hospital, your team performs monthly random audits on 100 claims. You classify a claim as defective or nonconforming if you find a mistake in any of the following categories:

1. Diagnosis-related group (DRG)
2. Primary insurance name
3. Attending physician name

So, the process has three opportunities for a defect per claim. Last year, your team audited 1,200 claims and found 60 mistakes in the DRG category, 36 mistakes in the primary insurance names category, and 84 mistakes in the attending physician names category. Overall, there are $60 + 36 + 84 = 180$ total defects out of $3 \times 1,200 = 3,600$ opportunities, which means that the probability p of a mistake can be obtained by:

$$p = \frac{D}{TOP} = \frac{180}{3,600} = 0.05 \quad (2.23)$$

Using Excel, you can find the $6\sigma_M$ measure by:

$$=ABS(NORMSINV(0.05))+1.5 \quad (2.24)$$

$$= 3.145 \quad (2.25)$$

By Equation 2.12, the DPMO measure in this scenario is estimated by:

$$PPMO = 0.05 \times 1,000,000 = 50,000 \quad (2.26)$$

Example 2.6 (Yield)

Following from Example 2.5, let's assume that your team found defects on different claims (so no claim had more than one defect). From this assumption, you can estimate the process **yield (Y)** as follows:

$$Y = 1 - \frac{\text{Number of defective claims}}{\text{Total number of claims}} = 1 - \frac{180}{1,200} = 1 - 0.15 = 85\% \quad (2.27)$$

Since, in general, a claim can only have one DRG, one primary health insurance name, and one attending physician, you can also estimate the yield from each category this way:

$$\text{Yield of DRGs} = 1 - \frac{60}{1,200} = 95\% \quad (2.28)$$

$$\text{Yield of insurance names} = 1 - \frac{36}{1,200} = 97\% \quad (2.29)$$

$$\text{Yield of attending physician names} = 1 - \frac{84}{1,200} = 93\% \quad (2.30)$$

The product of these rates produces a measure known as the **Rolled Throughput Yield (RTY)**[53]. The RTY measure is normally obtained by multiplying yields from independent steps in the process. So, in this case, RTY is given by:

$$RTY = 0.95 \times 0.97 \times 0.93 \approx 85.7\%. \quad (2.31)$$

So, what can you conclude?

1. One conclusion is that up to 15% (1-85%) of the claims could potentially be rejected for having the wrong billing information. In practice, rejected claims create rework and increase the average cycle time of the claims process as a result. We recall that any rework is muda and potentially expensive. In this example, Metropolis Hospital may not only have to pay someone to fix the rejected claims but also may incur additional costs of financing receivables^a.
2. Another conclusion is that the RTY measure provides a high-level view of the overall claims process. At the sub-process level, one does not see the whole picture. Let's assume that the claims process herein involves three different departments: Health Information Management (HIM) that assigns DRGs, Registration that selects attending physicians' names, and Billing that chooses the primary insurance names. At the department level, the HIM manager may be happy with a yield of 95%, and so would be the Registration manager with a yield of 93%. The Billing manager may even be happier with a yield of 97%. But, observe that collectively, only about 85.7% of the claims are clean, and this is **not good**.

^a**Receivables:** most health care providers don't receive payments as soon as they render a medical service. The payment or reimbursement could take weeks depending on how soon the organization submits a clean claim and the turn-around time of the third-party payers. We refer to unpaid claims as receivables. It is not uncommon for health care providers to finance their receivables in return for operating cash [22].

DFSS

DFSS¹³ is a variation of the Six-Sigma strategy that involves designing products and services at the Six-Sigma level quality from the get-go. In other words, the process will have little to no need for rework or improvement since it was designed at the desired standard. It is possible to accomplish this goal by involving customers in the design process or by simply inquiring about the customers' needs via surveys and focused groups [43].

2.6 Flowcharts

Flowcharts are graphical tools that we generally deploy during the *Define* phase of DMAIC to model the chronology of steps in a process. The types of flowcharts that we consider here include the **basic flowchart**, the **swim-lane diagram**, the **workflow diagram**, and **value-stream maps**.

2.6.1 A Basic flowchart

A **basic flowchart** depicts the chronology of steps in the process. Typical geometric shapes that we utilize for this purpose include a *rectangle* to suggest processing, a *diamond* (or a rhombus) to indicate a decision node, and a *D-looking shape* to show delay. We connect sequential steps with an arrow to suggest the direction of the flow of work. We show an example of a basic flowchart in Figure 2.2.

2.6.2 A swim-lane diagram

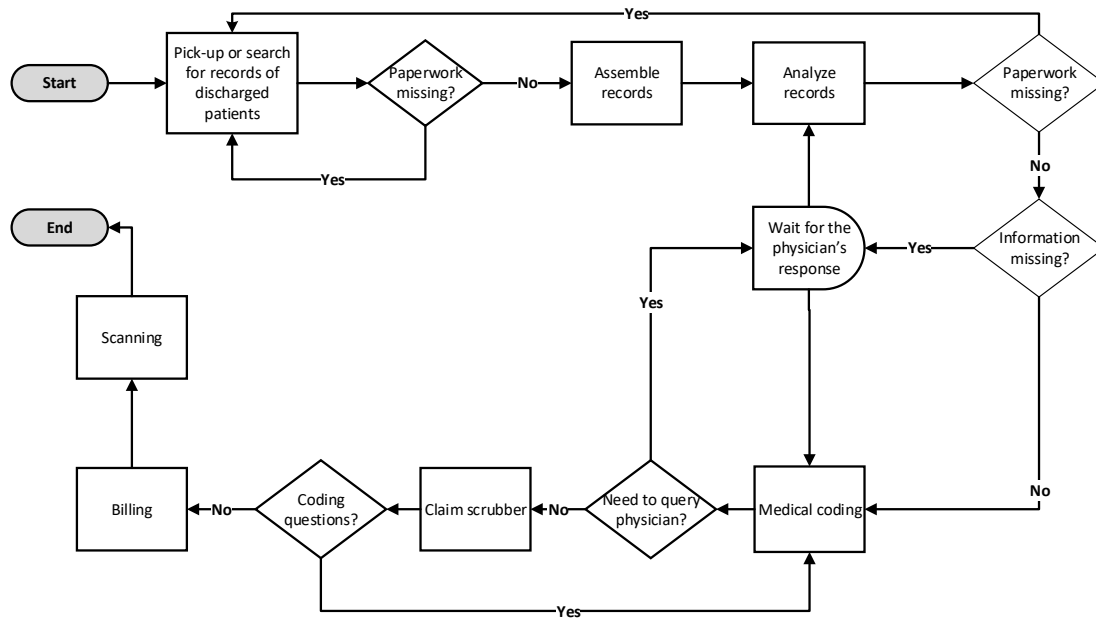
A **swim-lane diagram**, also known as a **cross-functional flowchart**, is similar to a basic flowchart, and we typically use the same geometric shapes. The only difference is that in the former, we attempt to group the steps by the function or the responsible person, whereas in the latter, we sequence the steps in the chronological order. Figure 2.3 shows an example of a swim-lane diagram that we adapted from the basic flowchart in Figure 2.2.

2.6.3 A workflow diagram

A workflow diagram is another type of flowchart that we consider. We use this diagram to illustrate the physical flow of work and transportation in a process. When drawing a

¹³DFSS: Design for Six Sigma

Figure 2.2: A basic flowchart for a discharge-to-billing process in a hospital that still uses paper records to document inpatient care services. All decision steps that suggest rework or going back to any of the previous steps should be examined for potential improvement.



workflow diagram, we attempt to represent the steps in the process using the shapes of physical objects (e.g., workstations, people, etc.). In Figure 2.4, we portray a type of a workflow diagram known as **spaghetti**¹⁴. From this chart, we appreciate that the denser the lines, the more traffic. For example, we can visually see that there is too much traffic between the **front-desk clerk** and the **printer**. Additionally, the workstations of **coding** and **analysis** seem to be physically far from the front-desk workstation, a setup that likely increases transportation. From the *Lean* strategy, all unnecessary transportation and movements are NVA [24]. In Figure 2.5, we portray a different type of a workflow diagram that indirectly relates to the same billing-to-discharge process that we initially illustrated in Figure 2.3. In this new representation, we attempt to portray the flow of work using physical objects that are typical in HIM operations. We draw the bolt-shaped lines to represent electronic communication. Regular straight lines imply the physical transfer of work.

2.6.4 Value-stream map

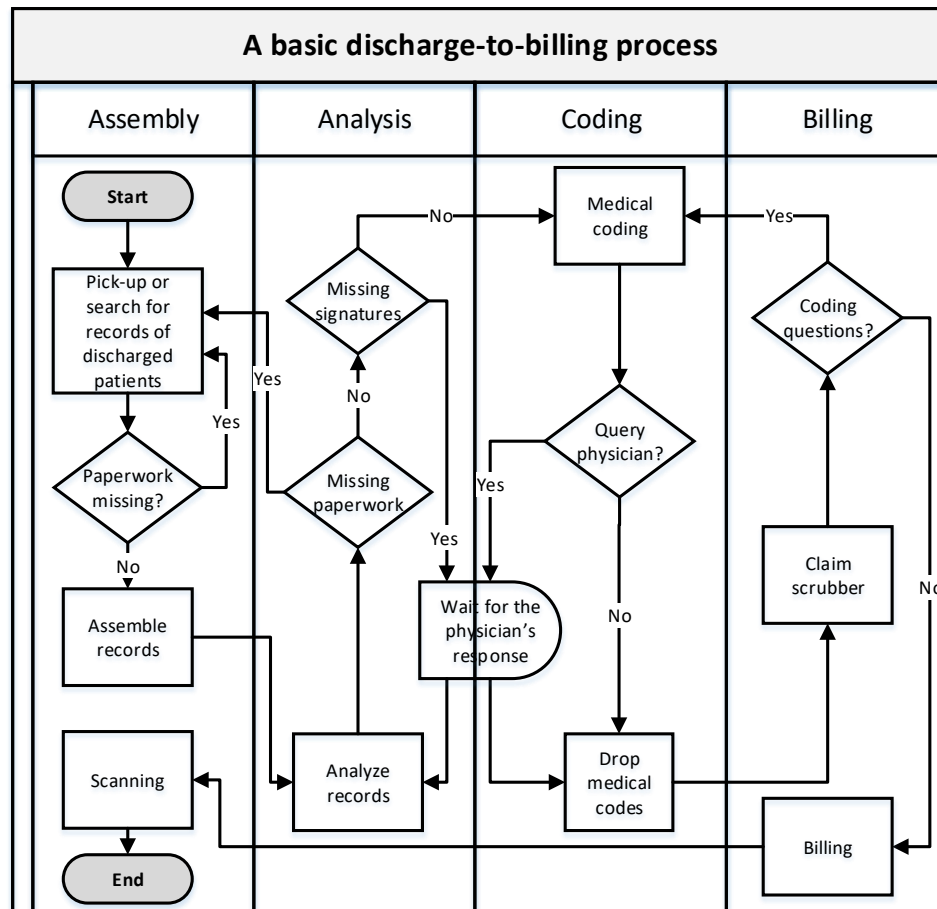
In a **value-stream map**, we model the sequence and the processing time (PT) of VA¹⁵ steps. All unnecessary waiting between two successive steps is considered muda¹⁶. We denote the overall cycle time as the lead time (LT). From this model, we can determine

¹⁴The name of **spaghetti diagram** comes from the observation that the lines in this diagram look like spaghetti noodles [14]

¹⁵**VA**: value-added

¹⁶**Muda**: a Japanese word used in Lean terminology to mean waste or NVA [53]

Figure 2.3: A swim-lane diagram adapted from the flowchart in Figure 2.2. As before, all steps that suggest rework are potential opportunities for improvement.



the PCE¹⁷ baseline measure this way [43]:

$$PCE = \frac{PT}{LT} \quad (2.32)$$

Figure 2.6 shows an example of a value-stream map for the same discharge-to-billing process under discussion.

¹⁷PCE: Process Cycle Efficiency

Figure 2.4: A Spaghetti diagram depicting the workflow related to the functions of assembly, analysis, and coding initially portrayed in the swim-lane plot in Figure 2.3

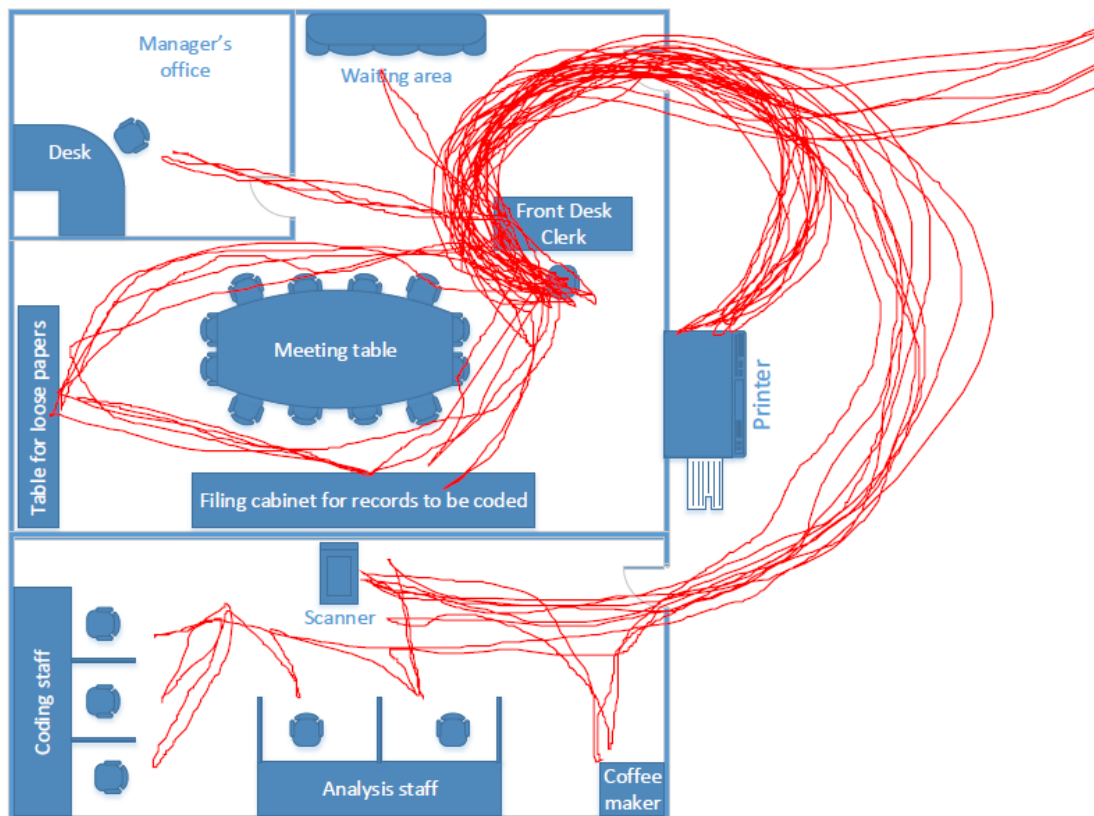


Figure 2.6: An example of a value-stream map for the discharge-to-billing process in Figure 2.2. The PCE measure is given by $PT/LT = 25.75/121.75 = 21\%$, which implies that about 80% of the discharge-to-billing cycle time is NVA. To improve the process, the lead time has to be decreased by removing all muda activities from the process. A good start would be the elimination of all causes of rework and unnecessary waiting.

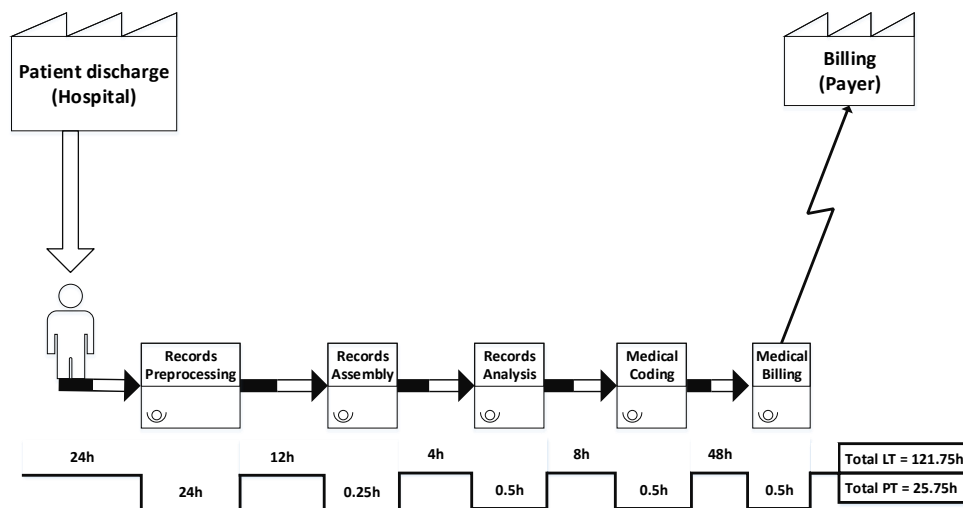
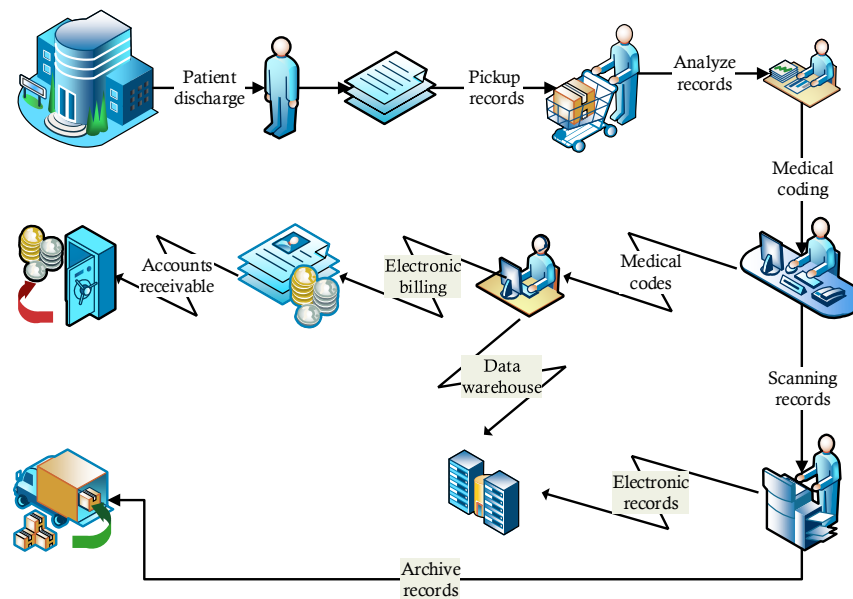


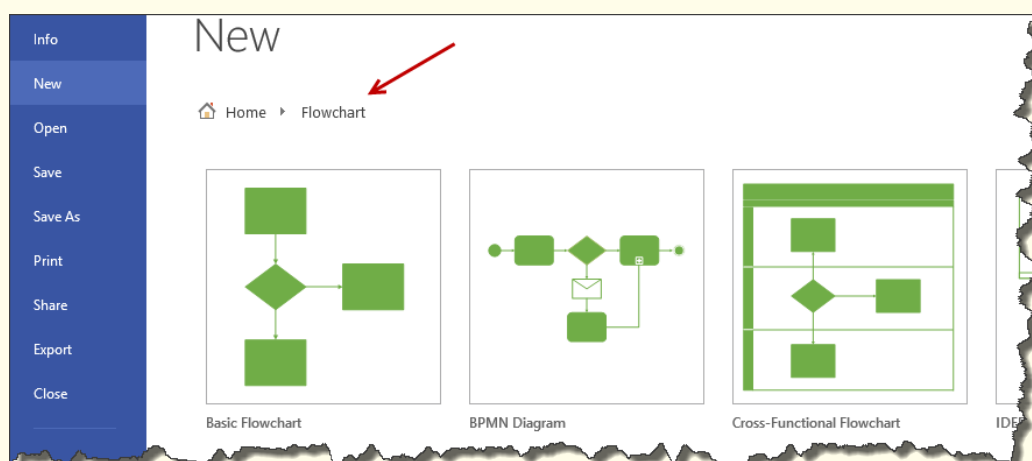
Figure 2.5: A workflow diagram related to the discharge-to-billing process that is depicted in Figure 2.2.



How-To 2.2 (Drawing flowcharts in Microsoft Visio 2013)

To draw flowcharts, click on *File > New > search for or click on Flowchart > choose the diagram of interest*. See a snapshot of Visio 2013 in Figure 2.7.

Figure 2.7: Flowchart options in Microsoft Visio 2013



To draw a floor plan for a spaghetti diagram, click on *File > New > search for Maps and Floor Plans*.

2.7 A project charter and a SIPOC diagram

After identifying opportunities for improvement, we have to devise an improvement strategy and make a business case defending the needs for investing scarce resources into our project. For this purpose, we employ tools such as a **project charter** and a **SIPOC**¹⁸ diagram to organize our project plan.

2.7.1 A project charter

We use a project charter to define an opportunity for improvement and to make a business case for the project. In this document, we also include the goal, the plan, the scope, and the team to carry out the project. A project goal should be **s**pecific, **m**easurable, **a**ttainable, **r**easonable, and **t**ime-sensitive (**SMART**) [50]. In Figure 2.8, we present an example of a project charter to make a case for improving quality in the discharge-to-billing process that we have previously plotted in Figures 2.2 - 2.6.

Figure 2.8: A project charter for improving the discharge-to-billing process in Figures 2.2 - 2.6

<p>Business Case This project is consistent with the goal of the finance department to reduce account receivables by 25% in the next 12 months.</p> <p>Goal Statement Increasing the PCE measure of the discharge-to-billing process from 21% to 75% over the next 12 months</p>	<p>Opportunity Statement Opportunity exists to streamline the discharge-to-billing process by eliminating NVA time related to medical query, analysis, and record preprocessing. The current PCE measure is 21%</p>
<p>Plan <i>Define:</i> 2 weeks <i>Measure:</i> 12 weeks <i>Analyze:</i> 6 weeks <i>Improve:</i> 12 weeks <i>Control:</i> 20 weeks</p> <p>Team Revenue cycle manager, billing manager, medical coding manager, quality specialist</p>	<p>Scope Medical coding and billing processes are mapped and the lead and processing times are measured using electronic time stamps in the hospital's main computer information system. NVA steps are reduced using Lean techniques. Variability is controlled using statistical process control tools</p>

¹⁸**SIPOC:** suppliers, inputs, process, output, customers [43]

Example 2.7 (A SMART goal)

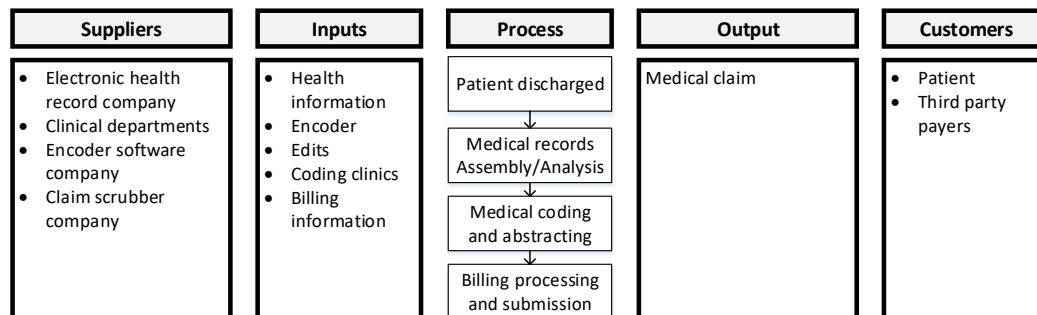
Working with the revenue cycle team, we will increase the PCE measure of the discharge-to-billing process from 21% to 75% over the next 12 months.

- S: Specifics** of the project are provided, including the focus, and who will be involved.
- M:** The outcome is **measurable** (21% to 75% of the PCE measure).
- A:** The stated goal seems **attainable** given the *time frame provided*.
- R:** The goal of achieving 75% of the PCE measure is **reasonable**. For example, the goal of reaching the PCE measure of 100% would be considered unreasonable.
- T:** The **time** frame of *12 months* is reasonable and is not overly too long or too short.

2.7.2 SIPOC diagram

We create a SIPOC diagram to illustrate a high-level view of the suppliers, inputs, process, output, and customer in the main process. In the **supplier** section, we list major suppliers of materials and information that we use in the process. In the **inputs** section, we document the materials and information that are critical for the process. In the **process** section, we display a general flowchart of major steps in the process. In the **output** section, we record outputs of the process that the customer cares about. In the **customer** section, we list relevant customers of the outputs of the process. Figure 2.9 portrays an example of a SIPOC diagram related to the project charter in Figure 2.8.

Figure 2.9: An example of a SIPOC diagram for improving the discharge-to-billing process in Figures 2.2 - 2.6



2.8 Check your understanding

1. Give an example of two techniques that we utilize to identify opportunities for quality improvement in a process?
2. How does **VOC** relate to **DFSS**?
3. What are the typical activities of **VOC**?
4. A **flowchart** that shows the chronology of cross-functional steps in a process is referred to as:
 - (a) a basic flowchart.
 - (b) a swim-lane diagram.
 - (c) a value-stream map.
 - (d) a spaghetti diagram.
5. We typically represent a **decision step in a flowchart** using which of the following geometric shapes?
 - (a) ◇
 - (b) △
 - (c) □
 - (d) D
6. Which of the following **flowcharts** tell us something about the physical workspace?
 - (a) A basic flowchart
 - (b) A swim-lane diagram
 - (c) A value-stream map
 - (d) A spaghetti diagram
7. What software would you recommend to draw a **swim-lane diagram**?
 - (a) Excel
 - (b) Visio
 - (c) Minitab
 - (d) PowerPoint
8. We can classify **the time between any successive VA steps** of a value-stream map as
 - (a) Non-value-added
 - (b) Value-added

- (c) Lead time
 - (d) Processing time
9. **Data profiling** is likely to be an activity that we perform during which of the following activities:
- (a) Customer surveys
 - (b) Quality audits
 - (c) Process mappings
 - (d) PCE calculations
10. Where would you classify a Xerox copier on a SIPOC diagram?
- (a) S
 - (b) I
 - (c) P
 - (d) O
 - (e) C
11. What happens to the process cycle efficiency when WIP is decreased?
12. How is a process that operates on a Push, instead of a Pull, basis likely to measure on the efficiency scale? Explain.
13. How did Motorola define the concept of Six-Sigma quality?
14. Given that VA time = 10 hours and NVA time = 20 hours, what is the estimated cycle time?
15. Given that VA time = 10 hours, NVA time = 20 hours, and the completion rate = 2 units per hour, what is WIP in this process?
16. A process has four independent steps, each having a yield of 95%. What is the RTY measure of this process? Interpret your results.
17. The probability of a defect in a process is 0.04. Estimate DPMO and σ_M rounded to the nearest hundredth.
18. The integration of information technology (IT) in clinical care has many benefits, including streamlined workflows and efficient clinical decision making [12]. But, the increase in the use of IT in clinical care has also come with significant managerial challenges related to hardware, software, and network issues. Indeed, some patient adverse events such as injuries and near misses have been attributed to IT processes [33]. As the new chief information officer (CIO) at Metropolis Hospital, create a hypothetical project charter to reduce IT-related adverse by 25% over a reasonable amount of time. Also, create a corresponding SIPOC. You can consult external resources as necessary.

19. Recreate the swim-lane diagram in Figure 2.3. In your new diagram, create a separate lane for the physician. Feel free to add more steps to the process.
20. Redesign the office layout in Figure 2.4 to minimize transportation. Why is your new layout better than the initial design? Explain your rationale in terms of the PCE measure.

MODULE 3

Tools for collecting and presenting the data

In this module, we review tools for collecting and aggregating data for quality improvement. The tools we consider include checksheets, pie charts, bar charts, Pareto charts, radar charts, and defect concentration diagrams. We implement these tools using Microsoft Excel and Minitab.

Key concepts and tools: *Data needs; Data collection; Data presentation; Pivot tables; Checksheets; Pie charts, Bar charts; Pareto charts; Radar charts; Defect concentration diagram*

Major objectives

After studying this module, you will be able to:

- 1. Define key concepts and tools for collecting and presenting the data*
- 2. Identify the type of data to collect*
- 3. Apply Pivot Tables in Excel*
- 4. Implement checksheets*
- 5. Implement pie charts*
- 6. Implement bar charts*
- 7. Implement Pareto charts*
- 8. Implement radar charts*
- 9. Implement defect concentration diagrams*
- 10. Survey other data collection and presentation tools*

3.1 Introduction

After we have defined an opportunity for improvement, we proceed to the *Measure* phase of DMAIC to collect data about the key inputs and output variables. We can query historical data if available. Otherwise, we need to gather new data.

3.2 The need for new data

During a quality improvement project, we may have to collect new data if [24]:

1. The required data have never been collected before.
2. The necessary data exist, but they do not answer the right question.
3. The needed data are available but not accessible.
4. The required data are available, but not well-organized.

One natural approach to collecting new data is **work sampling**, where we observe a process at random times and record performance indicators of interest [43]. Typically, we need at least 25 samples to determine the stability of the process [7, 43].

3.3 The types of data to collect

The types of data we collect depend on the nature of the project. At the minimum, the data we collect should allow us to determine the following **baseline measures** [24]:

1. **WIP**¹ to quantify the average amount of work in the system yet to be completed.
2. **The average completion rate** to capture the average number of items or service units that are completed per given time interval.
3. **The lead time** or cycle time to obtain the average amount of time that an item or a service unit spends in the system.
4. **The first-pass yield** to measure the percent of work that makes it through the system without rework.
5. **Time trap events** to help identify possible NVA activities such as the distance from the last step, the wait time, the number of approvals and handoffs, the setup time, the levels of learning curves, and reworks activities [24].
6. **Defects count** to help quantify any aspect of a product or service that is not conforming to the standard or the expectation of the customer.
7. **Descriptive statistics** such as the mean, median, variance, and standard deviation to gauge central tendencies and variability in the process.

¹**WIP:** Work-in-process

3.4 Tools for collecting and organizing data

Spreadsheet software, such as Excel, are effective mediums for collecting and storing the process data. To facilitate the application of analytic tools such as Pivot Tables, we should organize our data in the **first normal form** by eliminating repeating groups or multivalued attributes, and ensuring that each row is unique [28]. We exemplify how to create a first normal form in How-To 3.1.



How-To 3.1 (The first normal form) Table 3.1 depicts data that are not in the *first normal form* since attributes have repeating groups. For example, we can see that the ticket number, the location, and dates are omitted when their values are similar to those recorded in the previous row. One major problem with storing data in this manner is that we cannot easily aggregate, filter, or sort them.

Table 3.1: An example of data with repeating groups

Ticket#	Location	Date	Event
1001	PACU	1/2/2018	Software Data storage/backup User Interface Computer virus
1002	ER	1/17/2018	Hardware Network/server
1003	HIM	2/9/2018	System Configuration Hardware Record migration

In Table 3.2, we demonstrate how to reorganize the dataset in Table 3.1 into the first normal form.

Table 3.2: Data with no repeating groups (*first normal form*)

Ticket#	Location	Date	Event
1001	PACU	1/2/2018	Software
1001	PACU	1/2/2018	Data storage/backup
1001	PACU	1/2/2018	User Interface
1001	PACU	1/2/2018	Computer virus
1002	ER	1/17/2018	Hardware
1002	ER	1/17/2018	Network/server
1003	HIM	2/9/2018	System Configuration
1003	HIM	2/9/2018	Hardware
1003	HIM	2/9/2018	Record migration



Remark: Continuous quality improvement efforts require continuous data collection. To the extent possible, we should attempt to automate the data collection process, and consider storing our data in a relational database, instead of spreadsheets, to minimize data redundancy and update anomalies [28].

3.5 Tools for presenting the data

In this section, we introduce essential graphical tools to help summarize and present data. The specific instruments we consider are checksheets, pie charts, bar charts, Pareto charts, concentration diagrams, and radar charts. We will review other tools, such as histograms and control charts, in later modules.

3.5.1 Checksheets

Checksheets are SPC² tools that we primarily use **to summarize defects by their frequency and the time of occurrence**. While it is fairly easy to create checksheets manually, we recommend using software such as Excel, as we exemplify next.

Example 3.1 (Checksheets using Excel)

As a health informatics intern at Metropolis Hospital, you were asked to create a checksheet of information technology (IT) events by month. To accomplish this task, you decided to use Pivot Tables in Excel to produce a summary shown in Figure 3.1

Figure 3.1: A checksheet of IT events by month, Metropolis Hospital, 2018

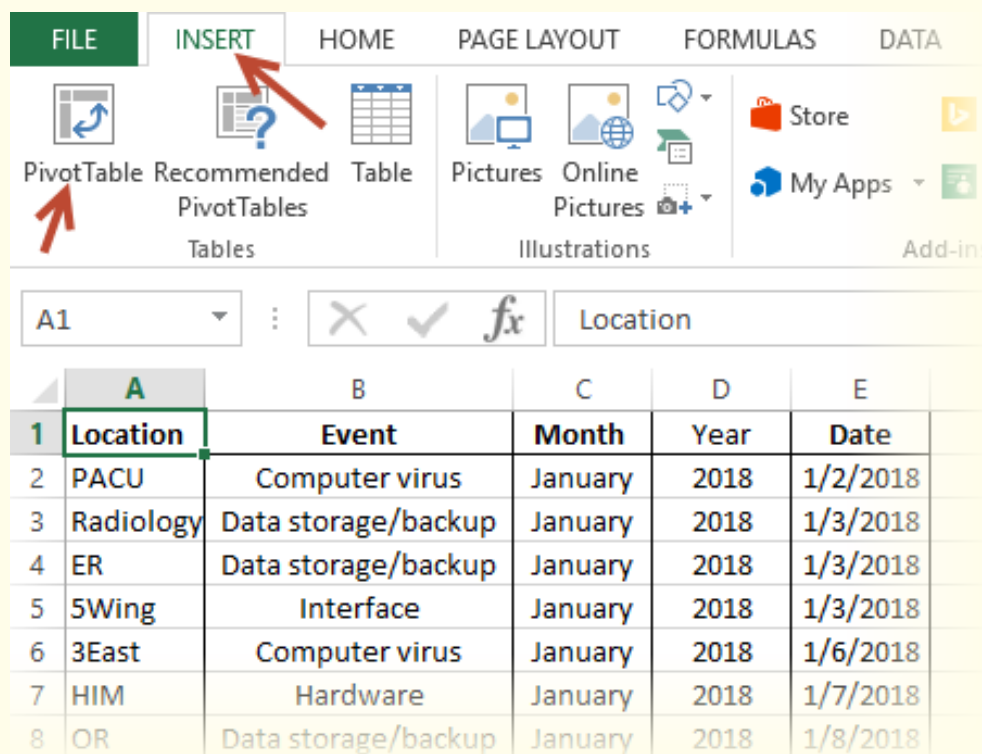
Count of Event Row Labels	Column Labels												Grand Total
	January	February	March	April	May	June	July	August	September	October	November	December	
Network/server	14	14	16	9	9	9	9	12	12	9	17	5	135
System Configuration	10	14	23	7	9	9	10	6	8	10	6	13	125
Data storage/backup	10	14	6	4	5	5	3	4	6	4	3	5	69
Computer virus	14	7	5	2	4	3	5	5	1	3		2	51
Interface	6	1		5	3	3	1		5	2	5	2	33
Record migration	1	7	4	2	1		2	3		5	1	3	29
Hardware	3		2	1	1			1				1	9
Software			1	1	1	2	1						6
Power failure						1		1				1	3
Grand Total	58	57	57	31	33	32	31	32	32	33	32	32	460

How-To 3.2 (Pivot Table in Excel 2013) To create a pivot table in Excel 2013:

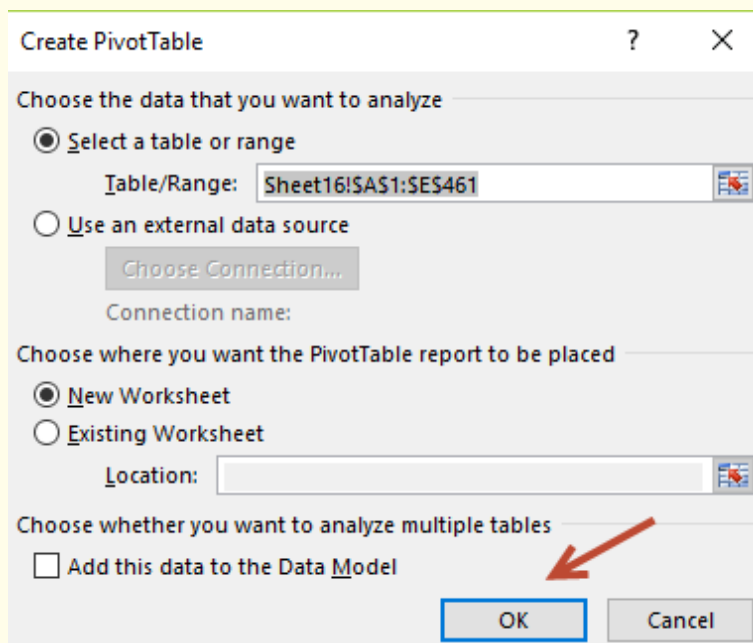


1. Click on *INSERT > Pivot Table* (see Figure 3.2)

²SPC: Statistical Process Control

Figure 3.2: Insert a Pivot Table in Excel 2013

2. Click OK on the *Create PivotTable* window that pops up (see Figure 3.3)

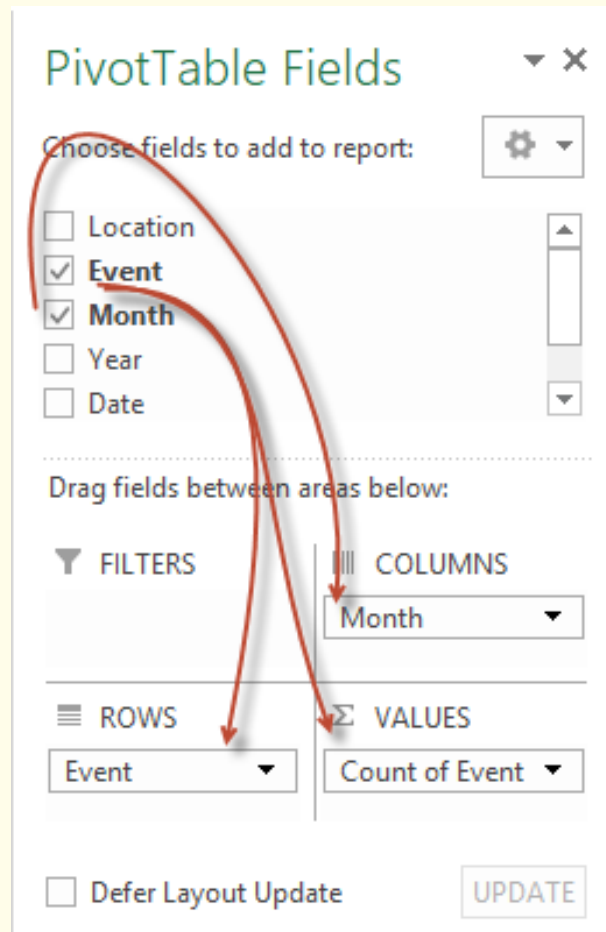
Figure 3.3: The *Create PivotTable* window in Excel 2013

3. Drag the fields of interest into the pan of *PivotTable Fields*. For example, to create the checksheet shown in Figure 3.1:

- (a) Drag and drop **Event** into **ROWS**
- (b) Drag and drop **Event** into **VALUES**
- (c) Drag and drop **Month** into **COLUMNS**

See the snapshot in Figure 3.4.

Figure 3.4: The pan of *PivotTable Fields* in Excel 2013



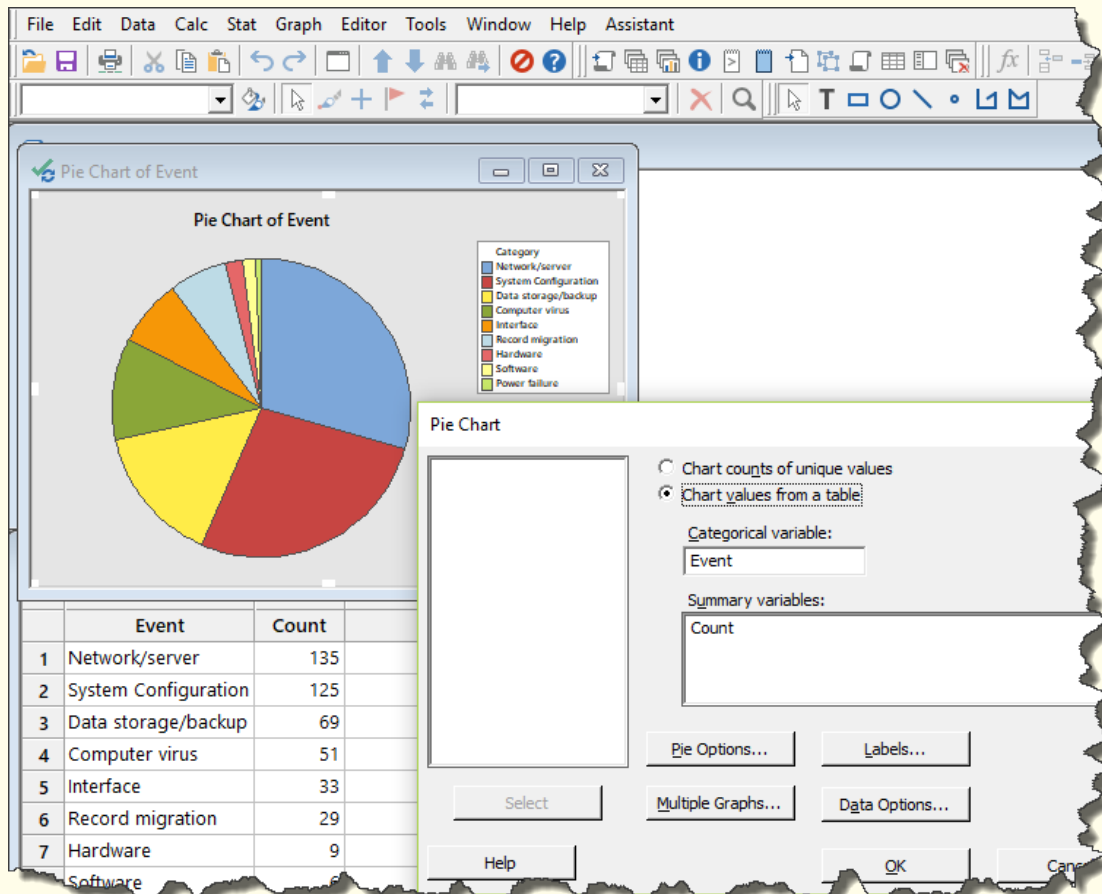
3.5.2 Pie Charts

Pie charts are graphical tools that we typically use to compare the proportions of different categories of the same attribute. To create a pie chart, we draw a circle and divide it into sectors that we call **pie slices**. The size of each slice is proportional to the share or count of the respective category. The bigger the slice, the more frequent the category.



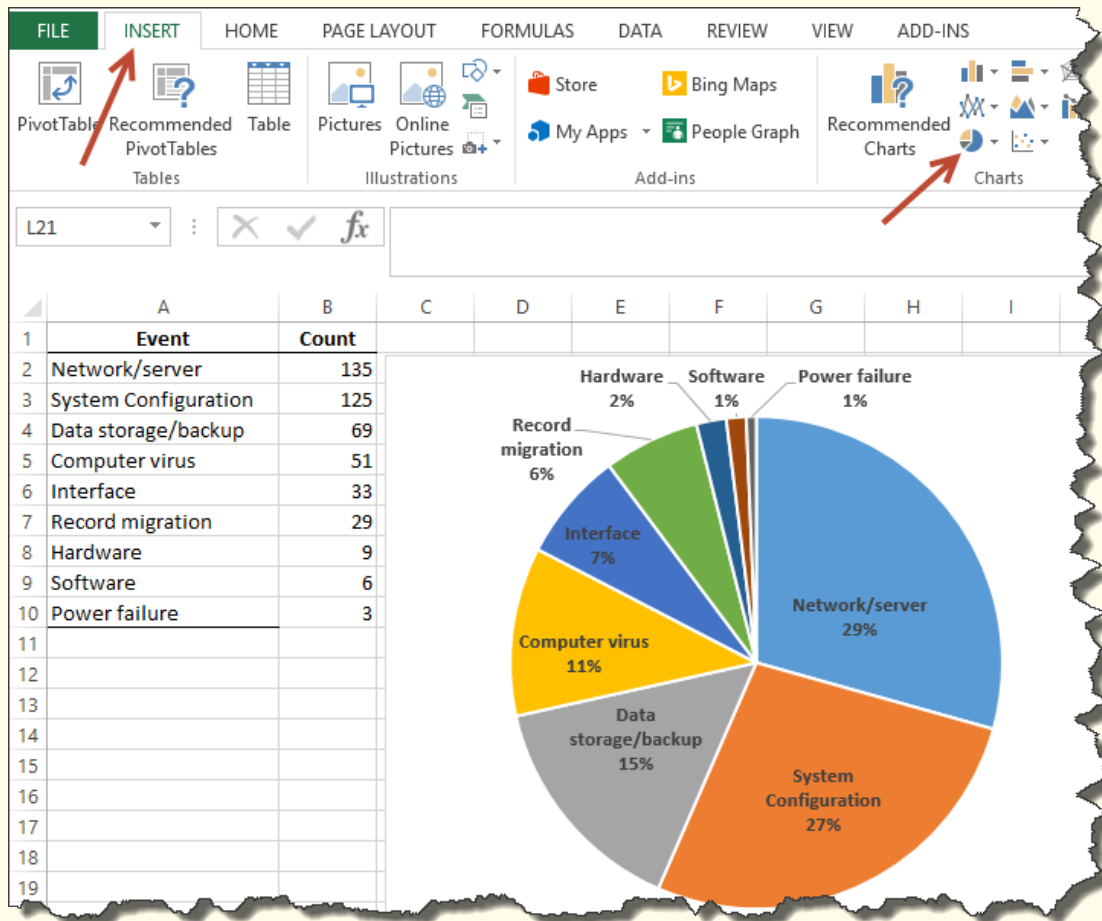
How-To 3.3 (Creating a pie chart in Minitab 18) To create a pie chart in Minitab 18, click on *Graph > Pie Chart > Chart values from table*. Upload your data and click OK. See the screenshot in Figure 3.5.

Figure 3.5: A screenshot of Minitab 18 with a pie chart based on the checksheet data in Figure 3.1



How-To 3.4 (Creating a pie chart in Excel 2013) To create a pie chart in Excel 2013, click on *INSERT > Pie Chart*. Several pie chart layouts are available. See an example in the screenshot depicted in Figure 3.6.

Figure 3.6: A snapshot of Excel 2013 with a pie chart based on the checksheet data in Figure 3.1.



The pie chart in this example reveals that the biggest slice is Network/Server (29%). The second biggest slice is System Configuration (27%), which is followed by Data storage/backup slice in the third place of frequency (15%). These three slices constitute about 71% of all IT events.

3.5.3 Bar charts

Bar charts are graphical tools that we use to model categorical and continuous data. We create bar charts by plotting a series of columns or rows of bars. The height or the width of each bar is proportional to the frequency of the corresponding category. In typical applications, we have to arrange bars in some logical order to allow for easy interpretation. One such arrangement is called Pareto, where bars are arranged in the descending order of frequency. For continuous data, we bin the data and arrange bars in the ascending order of the bin index. To show subcategories of major categories, we arrange bars in a **clustered** or **stacked** layout. For time series, we arrange bars in the

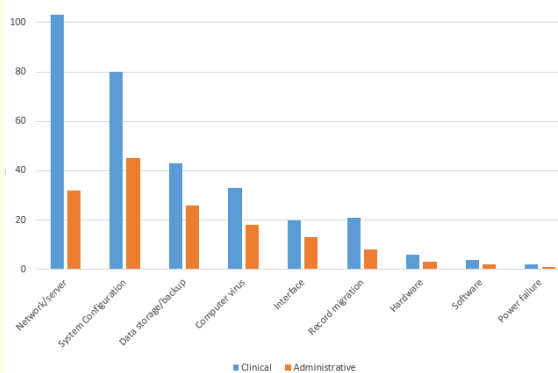
chronological order to time. Alternatively, we could use **line charts** to model time series.



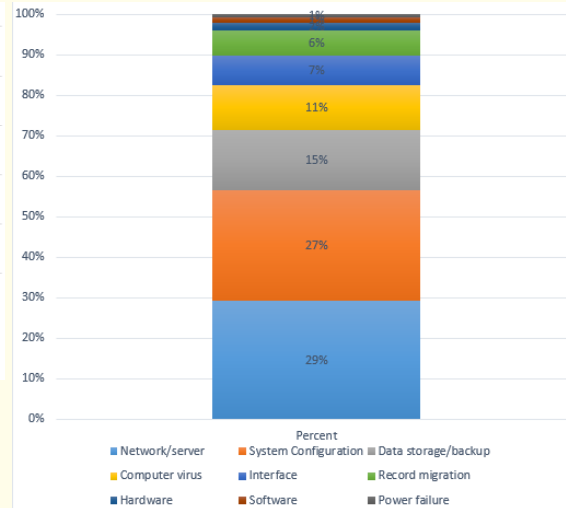
How-To 3.5 (Creating bar charts in Excel 2013) To create a bar chart in Excel, click on the *INSERT* tab > *Bar Charts* > *More Column Charts*. Several bar chart layouts are available. In Figure 3.7a, we show an example of a clustered bar chart. In Figure 3.7a, we portray a pie-like bar chart, which, in Excel 2013, is Layout 8. We show an example of a stacked bar chart in Figure 3.13.

Figure 3.7: Bar charts based on the data pictured in Figure 3.1

(a) A clustered column bar chart in Excel 2013



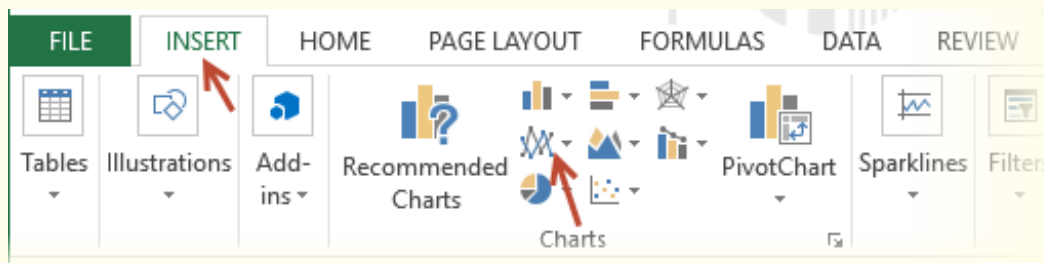
(b) Layout 8 of a bar chart in Excel 2013



How-To 3.6 (Line charts in Excel 2013)

To create a line chart in Excel 2013, *select your data* > *click on INSERT > Line charts*. See the snapshot in Figure 3.8.

Figure 3.8: Options for line charts in Excel 2013



3.5.4 Pareto charts

Pareto charts, also called *histograms for attribute data*, are primary tools of SPC that we use to isolate the *vital few* factors that cause the majority of defects [24]. We often

associate the **80/20 rule** with Pareto charts. This rule generally implies that 20% of the process steps cause 80% of defects. We can create a Pareto chart by first generating a bar chart and then arranging the blocks in descending order of the size of the bars.

How-To 3.7 (Creating a Pareto chart in Excel 2013)

To create a Pareto chart in Excel 2013:

1. First, determine the frequency of each category using Pivot Tables or the COUNTIF () function as follows:

- (a) Click on *Advanced Filter* as shown in Figure 3.9.

Figure 3.9: Advanced Filter in Excel 2013

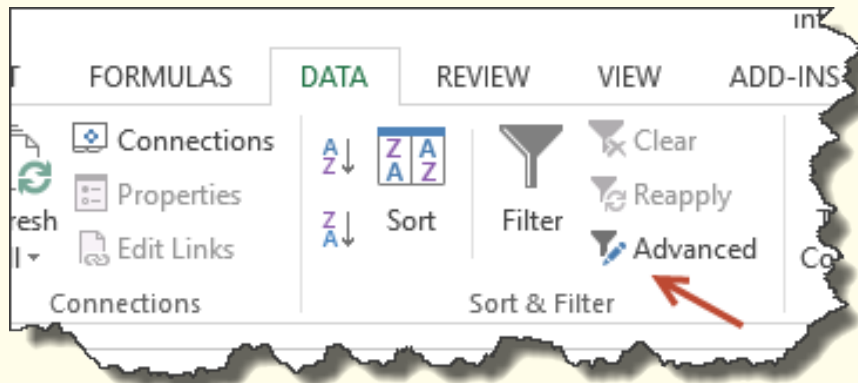


Figure 3.10: A snapshot of options of bar charts in Excel 2013



- (b) On the box that pops up,
 - i. Select a column for the *List range*,
 - ii. Click on *Copy to another location*,
 - iii. In *Copy to:*, choose a cell,
 - iv. Check the box of *Unique records only*.



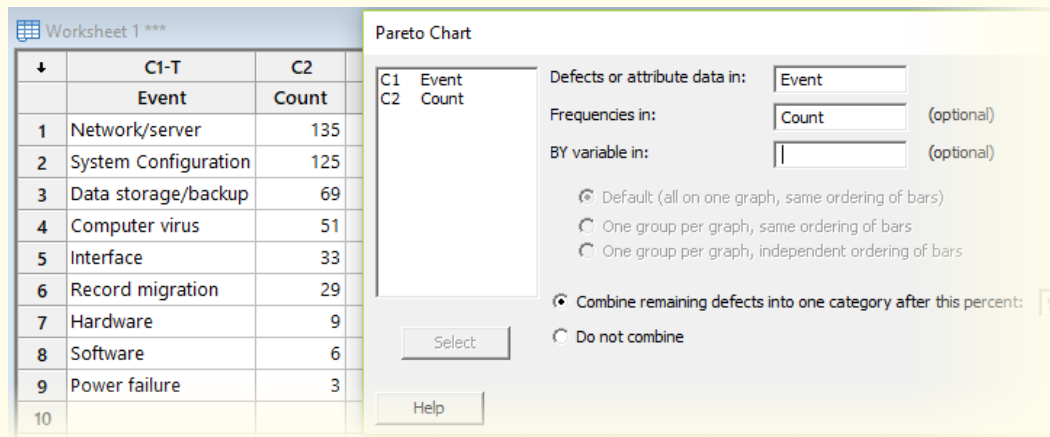
- (c) Now you can apply the COUNTIF() function to tally occurrences of each event (see Example 3.2).
- (d) Sort your results in descending order of frequency.

2. After sorting the frequency of categories, create a Pareto chart by inserting a bar chart or combo chart (see the snapshot in Figure 3.10).

How-To 3.8 (Creating a Pareto chart in Minitab 18)

To create a Pareto chart in Minitab 18, click on *Stat > Quality Tools > Pareto Chart*. In the *Defects or attribute data in*, select your categories. In the *Frequencies in*, select the count of your categories. See the screenshot in Figure 3.11.

Figure 3.11: Pareto chart in Minitab 18 using the checksheet data in Figure 3.1



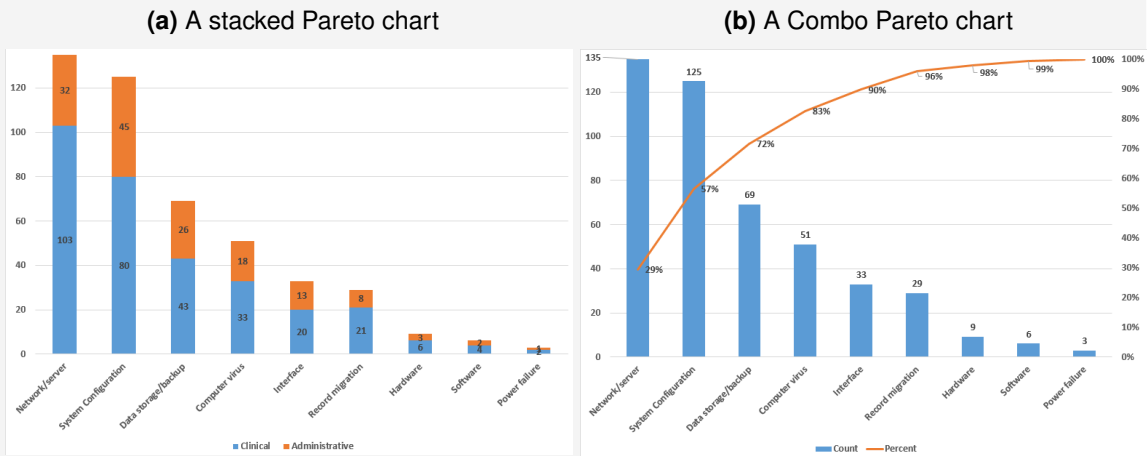
Example 3.2 (Pareto charts)

In Figure 3.12, we demonstrate how to apply the COUNTIF() function to generate the frequency of IT events.

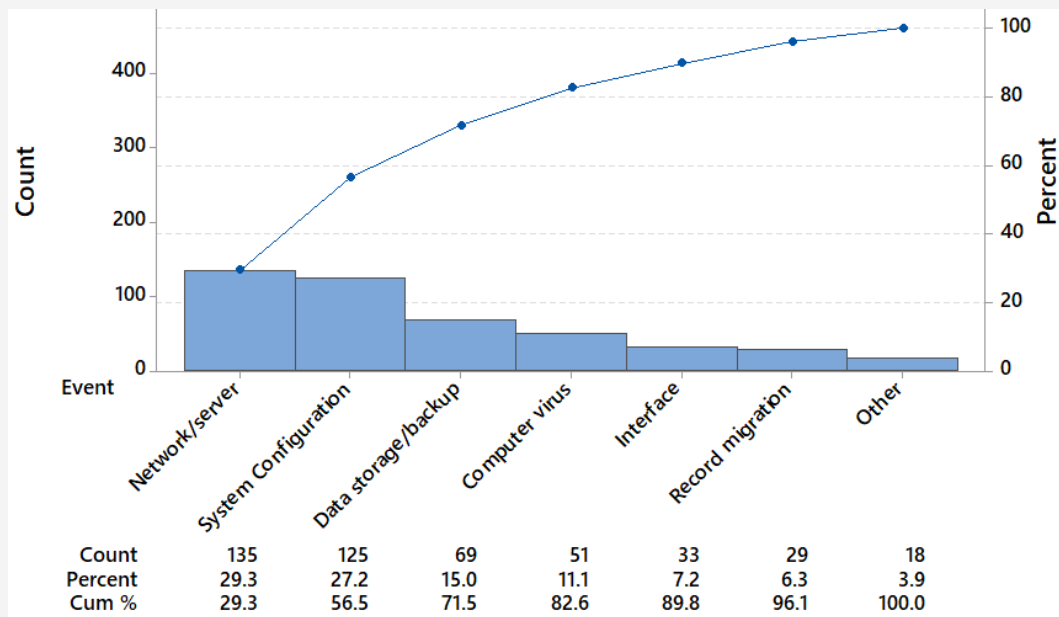
Figure 3.12: An example of how to apply the COUNTIF() function in Excel 2013

	A	B	C	D	E	F	G	H	I	J
1	Location	Event	Month	Year	Date		Event	Count		
2	PACU	Computer virus	January	2018	1/2/2018		Network/server	=COUNTIF(\$B\$2:\$B\$461,G2)		
3	Radiology	Data storage/backup	January	2018	1/3/2018		System Configuration	COUNTIF(range, criteria)		
4	ER	Data storage/backup	January	2018	1/3/2018		Data storage/backup	69		
5	5Wing	Interface	January	2018	1/3/2018		Computer virus	51		
6	3East	Computer virus	January	2018	1/6/2018		Interface	33		
7	HIM	Hardware	January	2018	1/7/2018		Record migration	29		
8	OR	Data storage/backup	January	2018	1/8/2018		Hardware	9		
9	er	Computer virus	January	2018	1/9/2018		Software	6		

After counting and sorting the data, we follow instructions in How-To 3.7 to create one Pareto chart using stacked bars (see Figure 3.13a) and another Pareto chart using a combo of bars and a line chart of cumulative percents (see Figure 3.13b).

Figure 3.13: Pareto charts created using Excel 2013.

From the cumulative percent line in the combo chart, we notice that the bars of Network/server, System Configuration, and Data storage/backup make up about 71% of the events. Figure 3.14 portrays a similar Pareto chart that we created using Minitab 18.

Figure 3.14: A Pareto chart created using Minitab 18 and the checksheet data in Figure 3.1

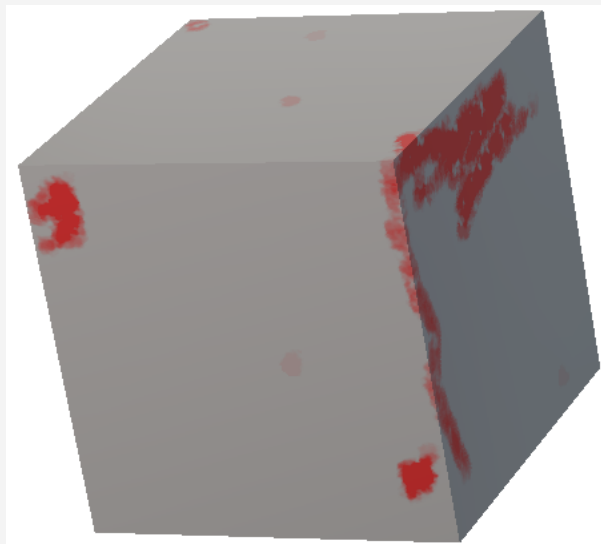
3.5.5 Defect concentration diagram

A **defect concentration diagram** is another primary tool of SPC that we commonly use to visually study the spatial distribution of defects on a physical product or model [43]. We can create this diagram by marking the locations of defects, as we demonstrate in the next example.

Example 3.3 (Defect concentration diagram)

We are interested in evaluating the quality of the cuboid widget in Figure 3.15 and make improvements as necessary. From internal inspections and customer complaints, we have marked locations of defects using red color, as shown in Figure 3.15. To improve this widget, we need to investigate all processes associated with the sites of defects and make appropriate improvements.

Figure 3.15: An example of a concentration diagram



3.5.6 Radar charts

Radar charts that we also call spider charts are graphical tools that we use to compare categorical data of the same attribute and scale. Structurally, radar charts look like spider webs with circular lines and axes, hence the name of spider charts. Each axis represents a distinct category.

How-To 3.9 (Creating a radar chart in Excel 2013)

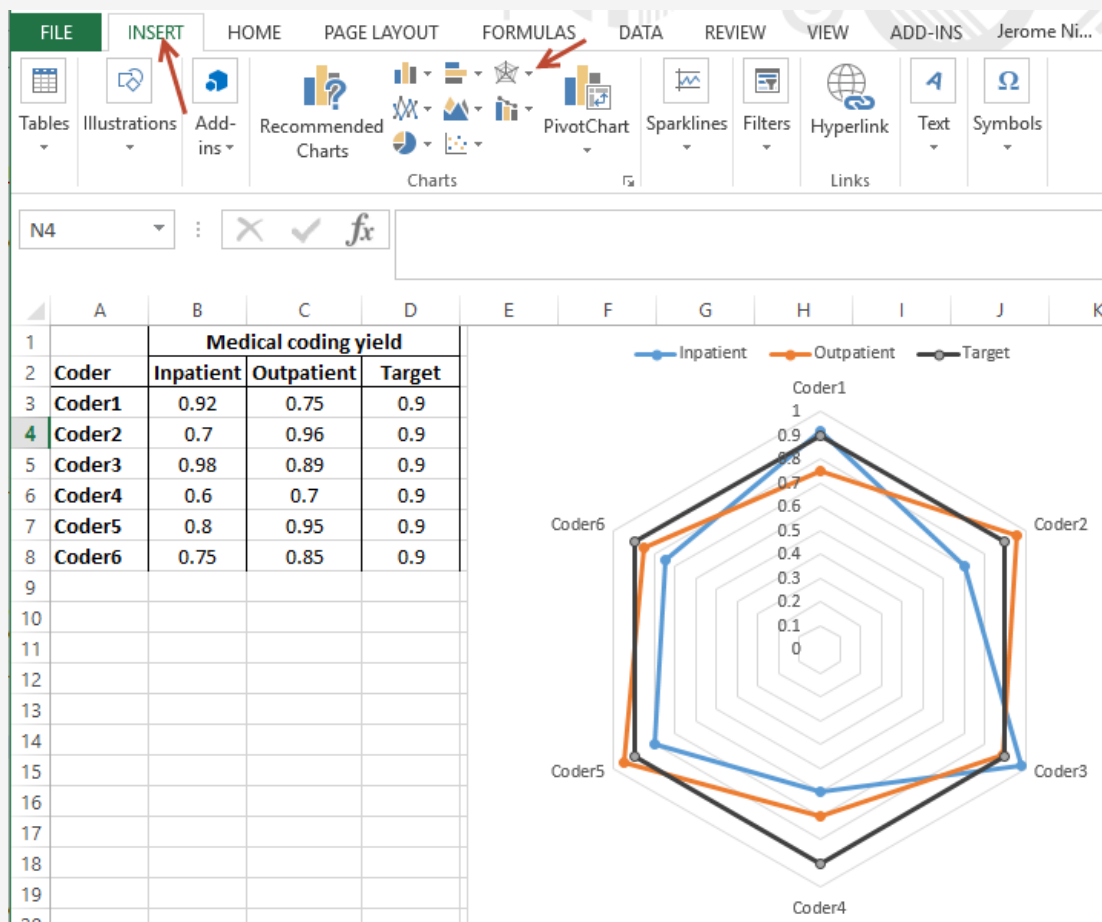
To create a radar chart in Excel, click on *INSERT > Radar chart*. See Example 3.4.

**Example 3.4 (A radar chart)**

A health informatics manager at Metropolis Hospital would like to evaluate the quality yield of coders' in the area of inpatient and outpatient. The target is both areas

is 90%. For this purpose, the manager created a radar chart portrayed in Figure 3.16.

Figure 3.16: An example of a radar chart comparing medical coding yield against the target



From this chart, we appreciate that Coder4 and Coder6 fail to meet the target in both areas of coding. Coder3 seems to be the most proficient in both areas of coding.

3.6 Check Your Understanding

1. When is it necessary **to collect new data** for a quality improvement project?
2. What does **work sampling** entail?
3. What are the benefits of organizing the data in the first normal form?
4. What are the time traps?
5. The amount of time that work spends in the system is captured using which of the following metrics?
 - (a) WIP
 - (b) Lead time
 - (c) First-pass yield
 - (d) Time trap
6. What advantages do **bar charts** have over **pie charts**?
7. We can create **checksheets** in Excel by using which menu option?
 - (a) PivotTable
 - (b) DataAnalysis add-in
 - (c) Insert Charts
 - (d) Create Checksheet
8. If you wanted to create **a summary of the count of defects** by month, using PivotTable in Excel, you would drag and drop — (check all that apply).
 - (a) the field of defects into ROWS
 - (b) the field of month into VALUES
 - (c) the field of defects into VALUES
 - (d) the field of the month into COLUMNS
 - (e) the field of the month into ROWS
 - (f) the field of defects into FILTERS
9. What is the ideal tool for **comparing the proportions** of categories in a dataset?
 - (a) Pie chart
 - (b) First normal form
 - (c) Defect concentration diagram
 - (d) Line chart

10. Which chart option in Excel is applicable when **constructing histograms**?
- (a) Bar chart
 - (b) Pie chart
 - (c) Scatter plot
 - (d) Radar chart
11. The 80/20 rule relates to which of the following charts:
- (a) Line chart
 - (b) Pie chart
 - (c) Pareto chart
 - (d) Radar chart
12. What is the best tool for representing the frequencies of data entry errors by **Department** and **Work-shift**?
- (a) Bar chart
 - (b) Pie chart
 - (c) Scatter plot
 - (d) Control chart
13. What tool looks **like a bar chart arranged in descending order** of frequency?
- (a) Pareto chart
 - (b) Run chart
 - (c) Control chart
 - (d) Cause-and-effect diagram
14. You would like to **compare the yield** of each employee against the target. What tools can you use for this purpose? (Select all that apply)
- (a) Pie chart
 - (b) Bar chart
 - (c) Radar chart
 - (d) Histogram
15. Which of the following tools is preferred for assessing the **spatial distribution** of defects on a product?
- (a) Defect concentration diagram
 - (b) Pie diagram
 - (c) Radar chart

(d) Checksheet

16. Use the given IT event dataset to create a checksheet.
17. Recreate the pie chart in Figure 3.6 using Excel or Minitab.
18. Recreate the Pareto chart in Figure 3.14 using Minitab. What events constitute about 80% of the count?
19. Recreate the radar chart in Figure 3.16. Using the same data, also create a corresponding stacked bar chart. Which data presentation chart do you prefer? Justify your answer
20. How would you go about converting the results from a defect concentration diagram into numerical values, and what would the benefits of doing so?

Tools for assessing the process variability and capability

In this module, we review statistical and graphical tools for assessing process variability and capability. Some of the graphical tools we discuss include histograms, stem-and-leaf diagrams, box plots, and control charts. We also demonstrate how to calculate statistical ratios for evaluating the capability of the process. For software, we use Minitab and Excel.

Key concepts and tools: Random sampling; Sample statistics; Variability; Probability distributions; Central limit theorem; Point estimators; Potential capability; Actual capability; PCR¹; Histograms; Stem-and-leaf diagrams; Box plots; Control charts

Major objectives

After studying this module, you will be able to:

- 1. Define key concepts and tools for assessing variability and capability*
- 2. Appraise the implication of variability on quality*
- 3. Compute sample statistics*
- 4. Explain the basic notion of the central limit theorem*
- 5. Master statistical measures of variability*
- 6. Create histograms*
- 7. Examine stem-and-leaf diagrams*
- 8. Implement box plots*
- 9. Analyze capability ratios*
- 10. Interpret probability plots*

¹PCR: Process capability ratio

4.1 Introduction

In this module, we discuss statistical and graphical tools for assessing variability and capability in the process. We recall that to **improve quality**, we must be able to evaluate and **reduce variability** in the process. Additionally, our process must be **capable** of meeting the specifications of the customer. The methods we consider include various statistical measures and graphical tools. For software, we use Minitab and Microsoft Excel.

4.2 Statistical background

4.2.1 Process sampling

Sample

In most quality improvement projects, it is not practical to observe and record everything that is happening in the process at all times. Even if such efforts were practical, they would likely be prohibitively expensive and time-consuming. A simplified approach is to use portions of the process to **infer** the performance of the entire process.

Definition 4.1 A **sample**, also referred to as **subgroup**, is a smaller set of measurements that we take from a larger set of the **population** measurements.

Terminology

We use the word **population** to suggest the entire process and the term **statistic** to indicate a statistical measure of the process, such as the **mean** (or the average) and the **variance**. The words **infer** and **inference** imply that we are making conclusions about the population using sample statistics. We will frequently use the expression **sampling distribution** to refer to the probability distribution of any sample statistic, most often the sample mean. We use letter n to denote the **sample size** (e.g., number of features) and letter m to symbolize the total number of samples that we call **sample number** (e.g., number of training examples).

Sampling techniques

Several sampling techniques apply to quality improvement, such as **random sampling**, where each item in the process has an equal chance of being selected. In **systematic sampling**, we set conditions for including an item in the sample (e.g., selecting every item in a prescribed position). In **stratified sampling**, we create distinct strata (e.g., groups) and choose an item from each stratum randomly or systematically. In **cluster sampling**, we group the population into naturally occurring clusters (e.g., geographical regions) and then sample from each cluster randomly or systematically. Both the stratified and cluster

sampling techniques tend to reduce bias by representing all relevant segments of the population [37].

Sampling musts

We must choose the proper **sample size** and the **frequency of sampling** to allow for more accurate inferential conclusions about our process. This requirement is particularly important for successful deployments of control charts to monitor the stability of the process. One rule of thumb is to take at least 25 samples before drawing any conclusions about the stability of a process [43]. In general, the larger the sample size, the more likely we are to detect unstable instances or out-of-control behaviors in the process. But, the practicality of large samples is not always feasible. Instead, we try to **take smaller samples, but more frequently** [43].

It is also imperative that when sampling, we ensure **rational subgroups** to promote the independence of the samples. By this, we imply that we should seek to maximize the difference between the samples while minimizing the difference within each sample. A natural approach to obtaining rational subgroups is sampling from **units produced during the same time interval** (e.g., same day). Another approach is **sampling only from units produced since the last subgroup** [43]. Without rational subgroups, erroneous conclusions about the stability of the process are likely, especially when samples were initially assumed to be independent.

4.2.2 Sample statistics

Typical sample statistics include the **mean** \bar{x} , the **variance** s^2 , and the **standard deviation** $s = \sqrt{s^2}$. Given a sample of size n , say (x_1, x_2, \dots, x_n) , we calculate the statistics \bar{x} , s^2 , and s , as follows:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (4.1)$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (4.2)$$

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (4.3)$$

We commonly refer to $\sum_{i=1}^n (x_i - \bar{x})^2$ as the **sum of squares (SS)** and $\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ as the **mean square (MS)**. Here, $n-1$ signifies the **degrees of freedom** or simply the number of independent values in a statistic [36]. We use the sample mean to estimate the average behavior of the process. Both the variance and standard deviation are non-negative measures of variability in the process. The bigger these quantities, the more variable the process and thus the inferior the quality. Unlike the variance which expresses squared variation measures, the standard deviation quantifies the process variability in the original units.



How-To 4.1 (Sample statistics in Excel 2013)

1. To calculate the sample mean, use the *AVERAGE()* function.
2. To calculate the sample variance, use the *VAR.S()* function.
3. To calculate the sample standard deviation, use the *STDEV.S()* function.

You can also calculate sample statistics via the *Data Analysis add-in > Descriptive Statistics > select your input range > check the option of Statistics summary > OK*.

If the Data Analysis add-in is not loaded in your Excel spreadsheet, you can add it by clicking on *File > Options > Add-ins > Analysis ToolPak > Manage: Excel Add-in > Check the box of Analysis ToolPak > OK*.



How-To 4.2 (Sample statistics in Minitab 18)

To calculate the sample mean, variance, and standard deviation in Minitab 18, click on *Stat > Basic Statistics > Display Descriptive Statistic > load your data into Variables > OK*.

Example 4.1 (Basic Statistics)

The manager of Metropolis Clinic has received several patient complaints about long waiting times. To investigate the matter, the manager performed a retrospective audit of 5 random visits each day for the last 25 days. The manager collected data in Excel, as displayed in Table 4.1.

From Table 4.1, we can deduce the following:

1. The **sample size** $n = 5$, since on each day, the manager selected five random visits.
2. The **sample number** $m = 25$ since overall, there are 25 samples.
3. The manager calculated the **sample mean** \bar{x} and the **sample variance** s^2 using Excel. For example, the manager calculated the mean of the sample on day # 1 as $= AVERAGE(B3 : F3)$, which equals 1.32. To determine the variance of this sample, the manager used $= VAR.S(\$B\$3 : \$F\$3)$, which resulted in 0.16075. To obtain the **standard deviation** of the same sample, the manager used $= STDEV.S(\$B\$3 : \$F\$3)$. Alternatively, the manager could have determined the standard deviation using $= SQRT(\$H\$3)$. By inserting the \$ sign in the formulas, the manager was able to populate the rest of the values by dragging down the formulas.
4. To ensure **rational subgroups**, the manager took one sample each day with the assumption that waiting times were relatively the same on each day, and daily visits were independent.

Table 4.1: A random sample of waiting times in hours, Metropolis Clinic

	A	B	C	D	E	F	G	H	I
1	Day #	Sample size					\bar{x}	s^2	s
2		1	2	3	4	5			
3	1	1.1	1.5	1.25	0.85	1.9	1.32	0.16	0.4
4	2	2.6	2.6	2.1	2.8	2.95	2.61	0.1	0.32
5	3	2.05	2.8	1.15	1	1.6	1.72	0.53	0.73
6	4	2.4	1.3	2.95	2.9	3	2.51	0.52	0.72
7	5	1.15	2.4	1.4	2.05	1.75	1.75	0.25	0.5
8	6	1.95	2.45	0.55	2.35	2.4	1.94	0.64	0.8
9	7	0.95	1.3	1.2	1.1	1.6	1.23	0.06	0.24
10	8	2.55	1.3	2.5	2.45	2.15	2.19	0.27	0.52
11	9	2.05	1.5	0.9	2.25	1.45	1.63	0.29	0.54
12	10	1.5	1.25	2.35	2	0.75	1.57	0.39	0.62
13	11	2	3	3	1.1	1.95	2.21	0.65	0.81
14	12	0.6	2.4	2.2	0.65	2.75	1.72	1.04	1.02
15	13	2	2.15	1.9	2.2	1.2	1.89	0.16	0.4
16	14	1.65	0.5	2.2	1.95	1.1	1.48	0.47	0.69
17	15	1	2.8	1.85	0.8	2.8	1.85	0.91	0.95
18	16	2.55	2.95	1.25	0.95	1.9	1.92	0.71	0.84
19	17	1.2	0.55	1.25	2.9	0.7	1.32	0.87	0.93
20	18	3	1.35	1.05	2.25	1.95	1.92	0.59	0.77
21	19	1	1.45	0.95	3	1.35	1.55	0.7	0.84
22	20	1.25	2.95	1.7	1.7	1.15	1.75	0.51	0.71
23	21	1.85	2.55	0.55	0.95	1.8	1.54	0.63	0.79
24	22	2.75	2.25	1	1	2.8	1.96	0.81	0.9
25	23	0.65	0.9	0.55	2.3	3	1.48	1.22	1.1
26	24	2.35	1.3	1.8	2.25	2.05	1.95	0.18	0.42
27	25	0.5	2.35	0.7	3	2.55	1.82	1.3	1.14
							$\bar{\bar{x}}$	$\bar{\bar{s}}^2$	$\bar{\bar{s}}$
							1.79	0.56	0.71

4.2.3 Probability distributions

Several probability distributions apply for quality improvement in health care informatics.

Definition 4.2 (Probability distribution)

A **probability distribution** is a mathematical function that relates each value of a random variable to a real number between 0 and 1 to express the relative chances of that value occurring.

In Box 4.1, we grouped the most common probability distributions into classes of **discrete** and **continuous**. We apply discrete distributions to model processes that generate random attribute data (e.g., number of data entry errors), whereas we use continuous distributions to model processes that generate random variable data (e.g., duration of system downtime).

Box 4.1 Common Probability Distributions

Type	Name	Probability Distribution	Mean $\mu =$	Variance $\sigma^2 =$
Discrete	<i>Bernoulli</i>	$p(x) = \begin{cases} p & x = 1 \\ 1 - p & x = 0 \end{cases}$	p	$p(1 - p)$
	<i>Binomial</i>	$p(x) = \binom{n}{x} p^x (1 - p)^{n-x}$ $x = 0, 1, \dots, n$	np	$np(1 - p)$
	<i>Geometric</i>	$p(x) = (1 - p)^{x-1} p$ $x = 1, 2, \dots$	$\frac{1}{p}$	$\frac{1 - p}{p^2}$
	<i>Poisson</i>	$p(x) = \frac{e^{-\lambda} \lambda^x}{x!}$ $x = 0, 1, \dots; \lambda > 0$	λ	λ
Continuous	<i>Normal</i>	$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2}$ $-\infty < x < \infty$ $-\infty < \mu < \infty, \sigma > 0$	μ	σ^2
	<i>Exponential</i>	$f(x) = \lambda e^{-\lambda x}$ $x \geq 0; \lambda \geq 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

It follows that for discrete distributions, $0 \leq p(x) \leq 1$ and $\sum_x p(x) = 1$. For continuous random variables, $0 \leq p(a \leq x \leq b) = \int_a^b f(x) dx \leq 1$ and $\int_{-\infty}^{\infty} f(x) dx = 1$.

Example 4.2 (Applying probability distributions to model data quality)

As a health care informatician at Metropolis Hospital, you are charged with assuring data quality in the electronic health record system. In your work, you classify a record as deficient when it has at least one data quality issue. If you were interested in modeling the probability of a deficient record, you would apply the Bernoulli random variable. If you were interested in the probability of the number of records with data quality issues, you would utilize the binomial random variable. You would use the Geometric random variable to model the number of discharges until a record has a data quality issue. To model the count of data quality issues per record, you would use the Poisson random variable. You would use the exponential random variable to model the time until a data quality problem occurs. To model the average amount of time required to fix a data quality issue, you would use the normal random variable. In all these scenarios, quality issues are assumed to occur independently. Additional assumptions may apply depending on the chosen distribution.

4.2.4 Normal distribution

During quality improvement projects, we tend to assume that our samples were drawn from a **normally distributed** process. This common assumption is the result of the central limit theorem (CLT).

CLT**Definition 4.3 (CLT and the sample mean)**

Regardless of the population probability distribution, the sampling distribution of the sample mean \bar{x} with size n , drawn from a population with mean μ and variance σ^2 , is about normally distributed and can be modeled as follows [43].

$$\bar{x} \sim Normal\left(\mu, \frac{\sigma^2}{n}\right) \quad (4.4)$$

It follows that as $n \rightarrow \infty$,

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim Normal(0, 1) \quad (4.5)$$

We commonly refer to the Z statistic in Definition 4.3 as the **Z-score**. The numerator of this statistic, $\bar{x} - \mu$, is known as the **sampling error** whereas its denominator, σ/\sqrt{n} , is referred to as the **standard error** [36]. By mapping Z into a probability space, as $\Phi(Z)$, we obtain the probability p of a given standard value being less than Z . To find the probability of a standard value being greater than Z , we calculate $1 - \Phi(Z)$. It follows that $Z = \Phi^{-1}(p)$.

**How-To 4.3 ($\Phi(Z)$ in Minitab 18)**

Store your Z-score value in a given column then click on *Calc > Probability Distributions > Normal > Input column > OK*.

How-To 4.4 ($\Phi(Z)$ in Excel 2013)

To calculate $\Phi(Z)$ in Excel 2013, use this cumulative distribution function: = *NORM.S.DIST(Z,1)*.

**Example 4.3 (Z-score)**

The expected mean and standard deviation of your process include $\mu = 30$ and $\sigma = 4$, respectively. If you drew a sample of size 10 records and found the mean to be 32, what would be the probability of observing a sample mean less than 32? To solve this problem, proceed as follows:

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{32 - 30}{4/\sqrt{10}} = 1.58 \quad (4.6)$$

You can find the probability of interest in Excel by = *NORM.S.DIST(1.581,1)* = 0.943. To calculate the probability of observing a sample mean greater than 32, use $1 - \Phi(Z) = 1 - 0.943 = 0.057$.

Sampling distributions related to Normal

Typical sampling distributions that relate to Normal are **Chi-square** (χ^2), **t-distribution**, and **F-distribution**. We characterize these distributions as follows:

1. Given independent and normally distributed sample data points x_1, x_2, \dots, x_n , where σ is known, $y = x_1^2 + x_2^2 + \dots + x_n^2$ is distributed according to χ^2 with $n - 1$ degrees of freedom (χ_{n-1}^2). We can express y as follows [43]:

$$y = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (4.7)$$

$$\equiv \frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2 \quad (4.8)$$

2. If σ is unknown, it follows that:

$$\frac{\bar{x} - \mu}{s/\sqrt{n}} \sim N(0,1) \left/ \sqrt{\frac{\chi_{n-1}^2}{n-1}} \right. \quad (4.9)$$

We refer to Equation 4.9 as the **t-distribution**, or **Student's t-distribution**, with $n - 1$ degrees of freedom [43]. We typically use this distribution to compare the means between two small samples, where we don't know the population standard deviation.

3. The ratio between two independent Chi-square random variables w and y with respective degrees of freedom u and v , is distributed according to **F-distribution**, as follows [43]:

$$F_{u,v} = \frac{w/u}{y/v} \quad (4.10)$$

We often use the F-distribution in the analysis of variance (ANOVA) to test the equality of the means from independent samples drawn from normally distributed populations with equal variances.

4.2.5 Probability plots

A probability plot is a visual statistical tool that we can employ to validate our assumptions regarding the probability distribution of our process. We can approximate a probability plot by fitting a line to data (e.g., data that fall between the 25th and the 75th percentiles) [43]. If the hypothesized probability distribution were reasonable, points would fall along the fitted line. We will often use this technique to verify the normality assumption.

How-To 4.5 (Normal probability plot in Excel 2013)

There is no dedicated option in Excel 2013 to create a probability plot unless one fits a regression model. But, we can approximate this plot manually, as we demonstrate in Example 4.4.



How-To 4.6 (Normal probability plot in Minitab 18)

Click on *Graph > Probability Plot > Single > Select your data > OK*. To change the type of distribution, click on the *Distribution* tab. See the snapshot in Figure 4.1.



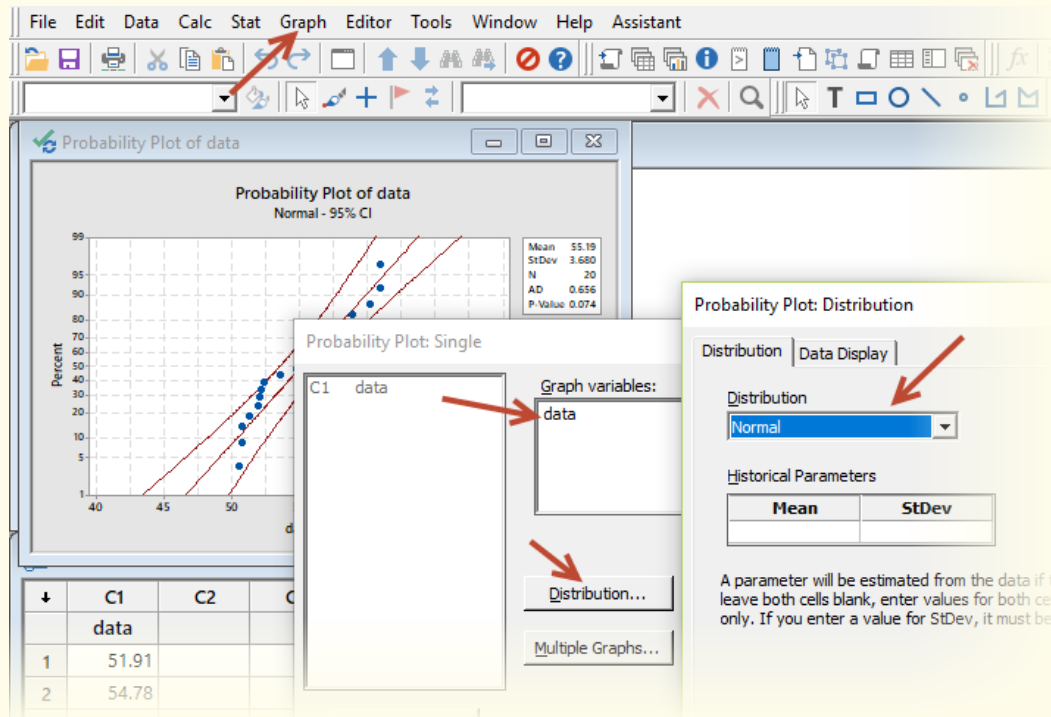
Figure 4.1: Options for a normal probability plot in Minitab 18**Example 4.4 (Estimating the normal probability plot in Excel 2013)**

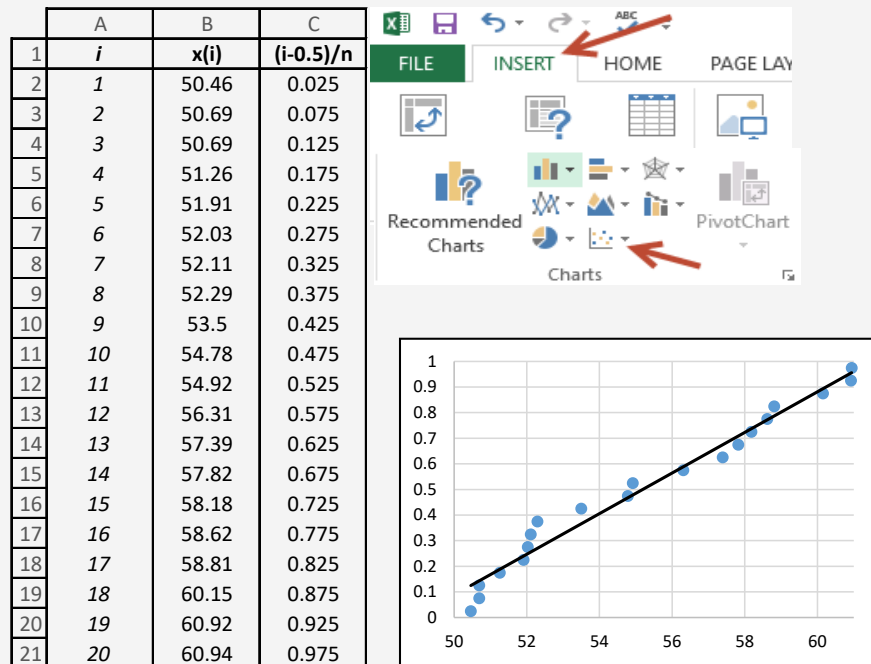
Table 4.2 presents random samples of operating room times of twenty random laparoscopic herniorrhaphy cases at Metropolis Hospital. The first column i enumerates the samples. The second column depicts operating room times in minutes $x_{(i)}$. The third column presents the cumulative frequencies given by $i - 0.5/n$. All columns are arranged in ascending order.

Table 4.2: Operating room times in minutes $x_{(i)}$, laparoscopic herniorrhaphy, Metropolis Hospital, January 2019

i	$x_{(i)}$	$(i - 0.5)/n$	i	$x_{(i)}$	$(i - 0.5)/n$
1	50.46	0.025	11	54.92	0.525
2	50.69	0.075	12	56.31	0.575
3	50.69	0.125	13	57.39	0.625
4	51.26	0.175	14	57.82	0.675
5	51.91	0.225	15	58.18	0.725
6	52.03	0.275	16	58.62	0.775
7	52.11	0.325	17	58.81	0.825
8	52.29	0.375	18	60.15	0.875
9	53.5	0.425	19	60.92	0.925
10	54.78	0.475	20	60.94	0.975

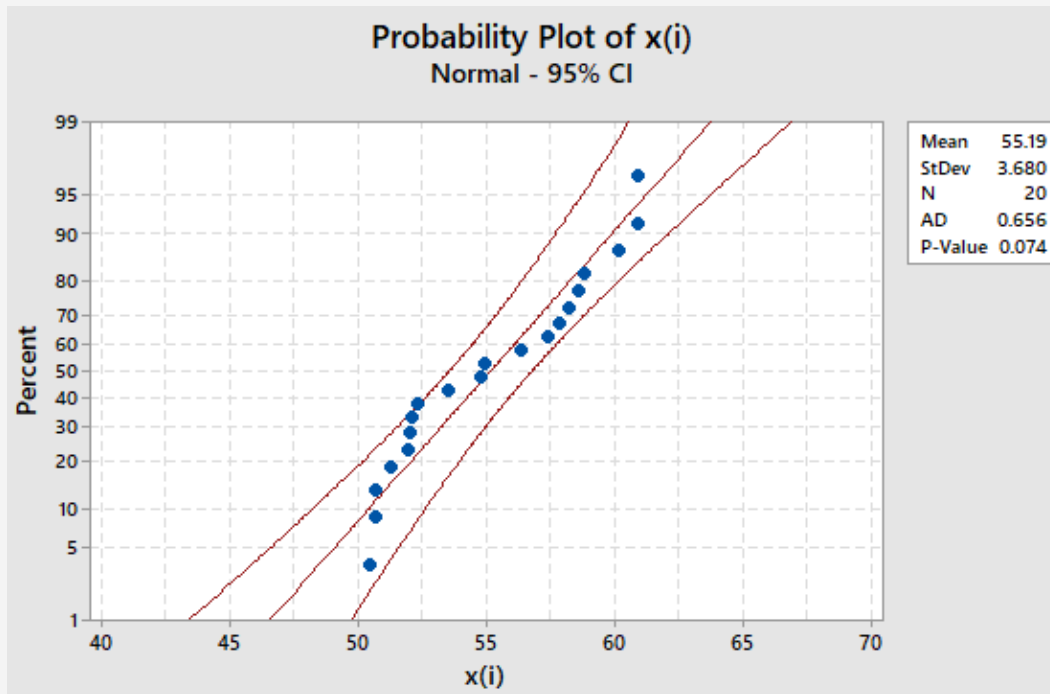
Figure 4.2 shows the setup of Excel for approximating a normal probability plot of the data in Table 4.2. The first step is **creating a scatter plot** between $x_{(i)}$ and $(i - 0.5)/n$. The second step is right-clicking on any data point and **adding a linear trend line**.

Figure 4.2: Approximating a normal probability plot of the data in Table 4.2



From the probability plot in Figure 4.2, we notice that the points tend to fall along the fitted line. Thus, we conclude that the data in Table 4.2 are about normally distributed.

We reproduced a similar plot using Minitab 18, as portrayed in Figure 4.3. Minitab results also show the **p-value** that we can use to make a conclusion about a hypothesis that the data in Table 4.2 are normally distributed. The given p-value is 0.074, which is greater than $1 - 95\% = 0.05$. Accordingly, we **fail to reject the null hypothesis** that the data in Table 4.2 are normally distributed at the 95% confidence level. We will discuss more about p-values in Module 7.

Figure 4.3: A normal probability plot of the data in Table 4.2 created using Minitab

4.2.6 Point estimators

In this subsection, we review the point estimation methods for the population mean and standard deviation parameters.

Definition 4.4 A **point estimator** is a statistic that we calculate from the process samples to estimate the population parameter [43].

The sample mean \bar{x} and variance s^2 are **unbiased point estimators** of the corresponding population mean μ and variance σ^2 since

$$E(\bar{x}) = \mu \quad (4.11)$$

$$E(s^2) = \sigma^2 \quad (4.12)$$

Here, E symbolizes expectation. But, the sample standard deviation is a **biased point estimator** of the population standard deviation σ , since

$$E(s) \neq \sigma \quad (4.13)$$

Continuous process

For continuous processes with the sample size $n > 10$, we can approximate the unbiased estimator of σ this way:

$$\sigma \approx \frac{\bar{s}}{c_4} \quad (4.14)$$

where \bar{s} is the mean of the sample standard deviations and c_4 is a constant that varies with n , as indicated in Appendix Table 1. When the sample size $n \leq 10$, we approximate the unbiased estimator of σ as follows:

$$\sigma \approx \frac{\bar{R}}{d_2} \quad (4.15)$$

where \bar{R} is the mean of sample ranges and d_2 is a constant that also varies with n , as indicated in Appendix Table 1 [43].

Discrete processes

For processes that generate attribute data, we can approximate the mean μ like this:

$$\mu \approx \bar{p} \quad (4.16)$$

where \bar{p} is the mean of the corresponding Bernoulli distribution. We apply the following approximation for the standard deviation σ :

$$\sigma \approx \hat{\sigma} = \sqrt{\frac{\bar{p}(1 - \bar{p})}{n}} \quad (4.17)$$

where n denotes sample size as before.

Example 4.5 (Point estimation)

This example is a follow-up to Example 4.1 and demonstrates how to calculate point estimators using the data in Table 4.1.

1. The manager determined the average of all sample means as $\bar{\bar{x}} = \frac{1}{25} \sum_{i=1}^{25} \bar{x}_i = 1.7932$. This value is the **unbiased estimator of the unknown process mean μ** . The manager obtained this value in Excel using `=AVERAGE(G3 : G27)`.
2. The manager took the average of all sample variances $\bar{s}^2 = \frac{1}{25} \sum_{i=1}^{25} \bar{s}_i^2 = 0.55877$. This value is the **unbiased estimator of the unknown process variance**. In Excel, the manager programmed this formula as `=AVERAGE(H3 : H27)`.
3. The manager obtained the average of the standard deviations like this: $\bar{s} = \frac{1}{25} \sum_{i=1}^{25} s_i$. In Excel, the manager programmed this formula as `AVERAGE(I3 : I27)`, which resulted in 0.708911687. To determine the **unbiased estimator of the unknown process standard deviation**, the manager

used Equation 4.14 as $\frac{\bar{s}}{c_4} = \frac{0.708911687}{0.94} = 0.75416137$. The value of c_4 was obtained from Appendix Table 1 when $n = 5$.

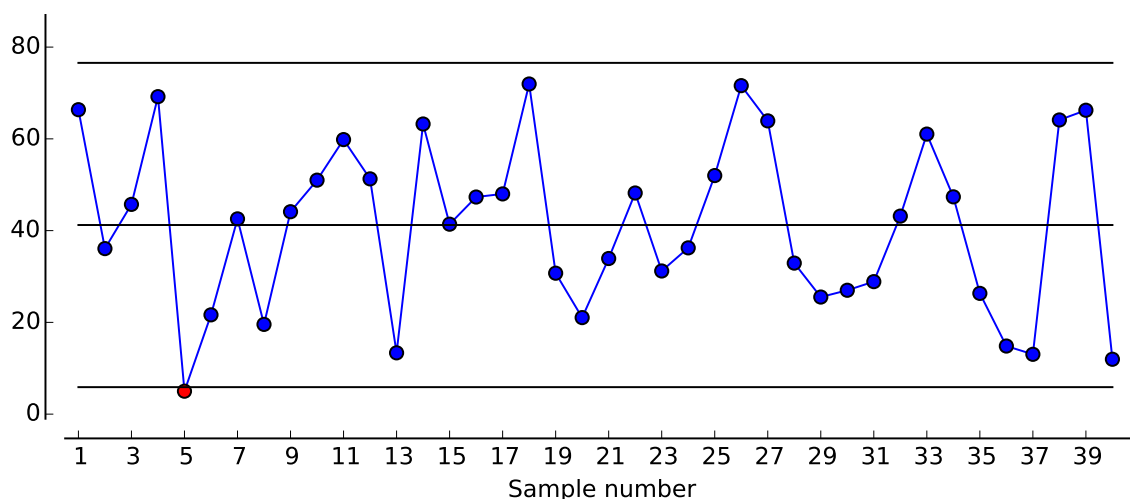
4.3 Process variability

In typical quality improvement projects, we use the sample variance or standard deviation to assess variability. If the quality metric of interest, say the process mean, is close to target, we can infer that the process has little variability. In other words, the process is stable. In contrast, a sample statistic that is further from the target is indicative of a process with high variability.

4.3.1 Control charts

We can conduct a hypothesis test of whether the process is stable. This test is implied in control charts when monitoring variability in the process. When a sample statistic falls outside of the control limits, we reject the null hypothesis that the process is stable. See an example of a control chart in Figure 4.4. As it is evident in this figure, sample 5 falls outside of the lower limit, which points to an unstable process. We will return the topics of hypothesis tests and control charts in later modules.

Figure 4.4: An example of a control chart



4.3.2 Histograms

A histogram is a type of bar chart that we use to assess variability in processes that generate continuous data.

Constructing a histogram

To construct a histogram, we first group data into **bins**, also called **intervals** or **classes**. We must choose an appropriate binning strategy for a histogram to be informative. Given the lack of a universal agreement about how to structure the bins, one common rule of thumb to select the number of bins k such that:

$$k = \left\lceil \sqrt{n} \right\rceil \quad (4.18)$$

where n is the number of observations in the data and the ceiling symbols $\lceil \cdot \rceil$ signify *rounding up*. Another formula that we can use follows from Sturges' rule and looks like this [43]:

$$k = 1 + \log_2 n \quad (4.19)$$

where \log is a mathematical symbol for logarithm. After deciding on the number of bins, we calculate the width of each bin h like this:

$$h = \frac{\max(x) - \min(x)}{k} \quad (4.20)$$

where x represents the sample data. Subsequently, we determine the range of the bins as follows:

$$bin_1 = [\min(x), \min(x) + h + 1) \quad (4.21)$$

$$bin_2 = [\min(x) + h + 1, \min(x) + 2h + 1) \quad (4.22)$$

...

$$bin_k = [\min(x) + (k - 1)h + 1, \min(x) + kh + 1) \quad (4.23)$$

Finally, we throw each observation into the bin with a suitable range, and construct a histogram by graphing a bar chart in ascending order of the bin index (e.g., bin_1, \dots, bin_k).

Interpreting a histogram

By visually examining the shape of a histogram, we can make inferences about the process variability regarding the skewness, kurtosis, modality, and the likely probability distribution. We can also learn about other characteristics of the process, such as the central tendency (e.g., mean or median) and outliers. We show examples of typical histograms in Figure 4.5.

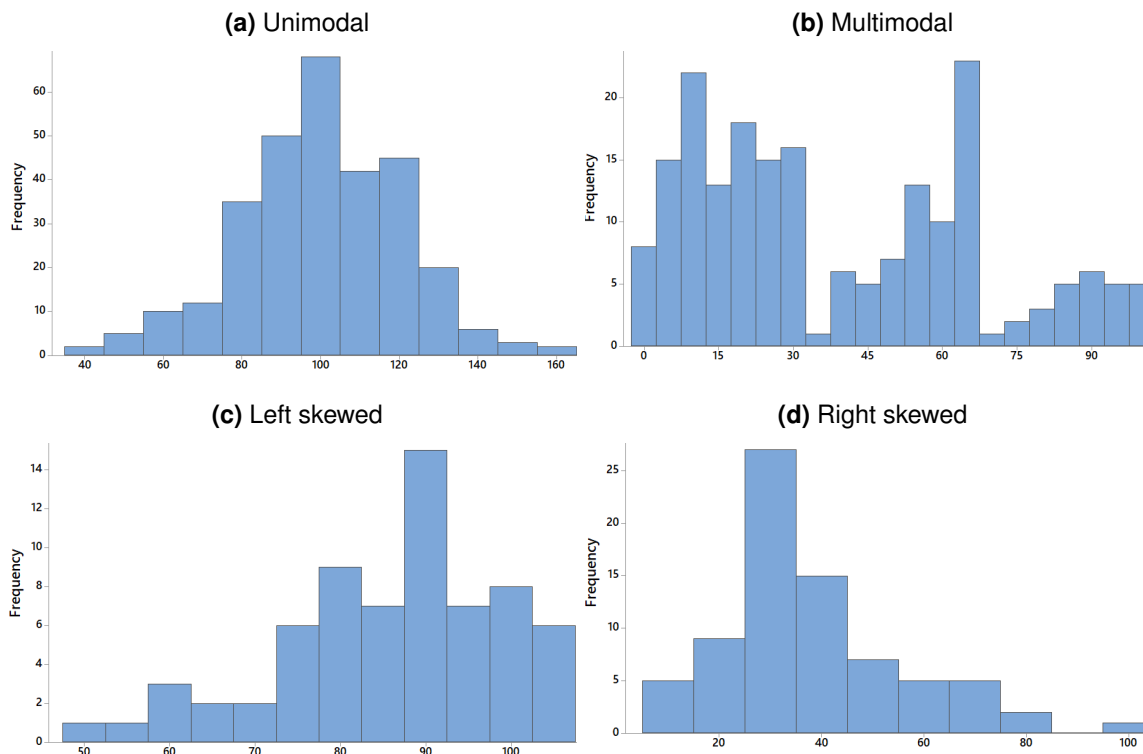
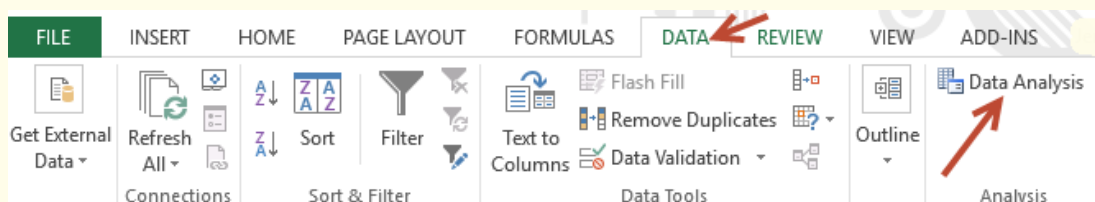
Figure 4.5: Examples of histograms

Figure 4.5a shows a bell-shaped unimodal (one peak) histogram suggesting a normal distribution with the mean around 100. Figure 4.5b shows a multimodal histogram that suggests a mixture of probability distributions. Figure 4.5c shows a histogram with a left-skewed distribution. Figure 4.5d displays a histogram with a right-skewed distribution and a possible outlier in bin 100. Histograms with fewer outliers will generally have a low kurtosis measure. The latter quantity describes the degree of the *peakedness* and *flatness* (e.g., heavy-tail or light-tail) of the given distribution as compared to the normal distribution [19].

How-To 4.7 (Creating a histogram in Excel 2013)

To construct a histogram in Excel, follow these steps:

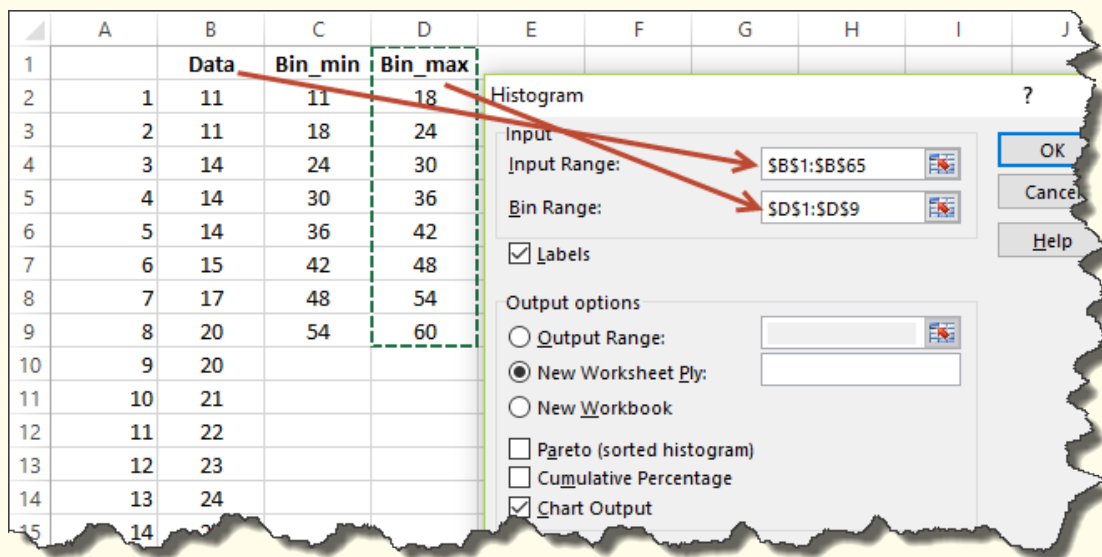
1. First, load the **Data Analysis Tools** add-in. In Excel 2013, you should see this add-in by clicking on the *Data* tab and looking under *Analysis*, as we illustrate in Figure 4.6.

**Figure 4.6:** Data Analysis add-in in Excel 2013

If the add-in is not loaded, you can load it by clicking on *File > Options > Add-ins > Analysis ToolPak > Manage: Excel Add-in > Check the box of Analysis ToolPak > OK*.

2. Input all your data into one column
3. Determine the range of each bin (see Equations 4.21 - 4.23).
4. Next, click on *Data Analysis > Histogram > select your sample data for **Input Range** > select *Bin_max* data for **Bin Range** > check **Chart Output***. See the snapshot in Figure 4.7.

Figure 4.7: Histogram options in Excel 2013



5. Instead of the *Data Analysis* add-in, you can also use the *COUNTIFS()* function to tally observation in each bin. For example, in reference to Figure 4.7, you can count the number of observations that fall in the range of cells C2 and D2, as follows: `=COUNTIFS(B1:B65,">="&C2,B1:B65,"<"&D2)`

How-To 4.8 (Creating a histogram in Minitab 18)

To construct a histogram in Minitab, follow these steps:

1. Click on *Graph > Histogram > Simple > Select your data > OK*.
2. To change the number of bins, double-click on any bar in the histogram > *Binning tab* > select the *Number of Intervals* > input the desired number > *Ok*.
3. To create a cumulative histogram, click on *Graph > Histogram > Simple > Select your data > Scale > Y-Scale Type > check **Percent and Accumulative values across bins** > OK*.

Remark: We recommend using Minitab to create histograms instead of Excel since the former gives the user more control to change the inputs. Minitab also allows for fitting distributions to data, an option that Excel 2013 does not have.



4.3.3 Stem-and-leaf diagrams

Just like histograms, we use stem-and-leaf diagrams to assess variability in continuous data.

Constructing a stem-and-leaf diagram

To create a stem-and-leaf diagram, we first generate the *stems* and then match the leading digit(s) of our observations to the appropriate stem. The remaining digits constitute *leaves* that we arrange in ascending order. We typically create stems using numbers between 0 and 9. For example, if we had an observation 76, **7** would be matched to the stem of 7 and **6** would be part of the leaf. Stems with no leaves are deleted. In a **full-stem**, all numbers with the same stem create one leaf. In a **half-stem**, two leaves result from a single stem. Half-stem-and-leaf diagrams tend to exhibit flatter distributions as compared to the diagrams of full-stems.

Example 4.6 (Half-stem vs. Full-stem) In a full-stem, observations 76, 70, 73, and 79 create one leaf as follows: **7 0369**. In a half-stem, we would have two leaves that split at 75. One leaf would look like this **7 03**, and the other would be as follows **7 69**.

Interpreting a stem-and-leaf diagram

Just like in histograms, the shape of a stem-and-leaf diagram informs us about the process central tendency, skewness, outliers, and the likely probability distribution. By the central tendency here, we are generally referring to the median or the 50th percentile of the data. To find the *position* of a given *percentile*, we use the following formula [36]:

$$position = (percentile) * (n + 1) \quad (4.24)$$

where n is the number of observations.

How-To 4.9 (Creating a stem-and-leaf diagram in Minitab 18)

To create a stem-and-leaf diagram in Minitab 18:

1. Click on *Graph > Stem-and-Leaf > select your data > OK*.
2. To change the number of stems, enter the increment number of 5 or 10 in the increment box. The increment of 10 is for creating full-stems, whereas the increment of 5 is for creating half-stems.



Example 4.7 (Stem-and-leaf diagram) The data in Table 4.3 shows a sample of days in accounts receivable at Metropolis Clinic. We will help the manager assess variability in this dataset.

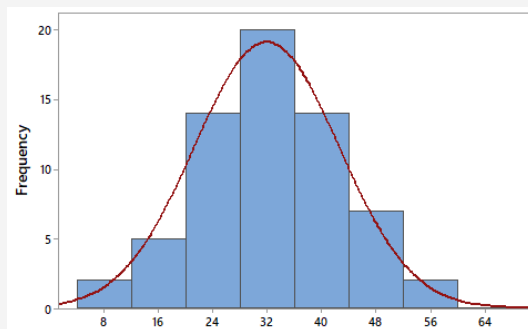
Table 4.3: Days in Patient Accounts Receivable at Metropolis Clinic

38	47	30	22	32	51	40	32	31	36	24	26	57	14	38	26
15	31	20	26	41	28	40	37	48	28	32	44	39	45	33	20
47	32	26	32	27	38	28	34	27	37	28	11	42	25	33	23
59	14	50	28	24	17	21	38	28	14	30	34	11	36	41	34

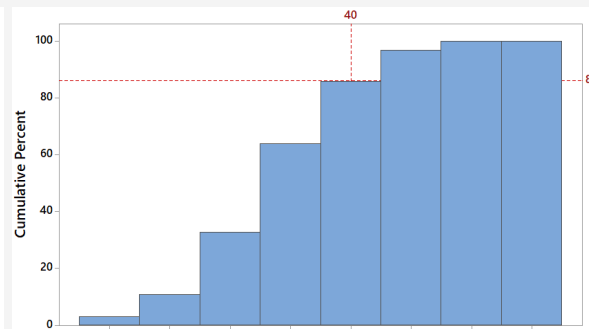
In Figure 4.8, we show two types of histograms that we created using Minitab 18. We used Equation 4.18 to change the number of bins to 8. Concerning Figure 4.8a, we can infer that the data are about normally distributed with the central tendency of about 32 days on average. We fitted a normal distribution to the data, which agrees with the apparent normality of the dataset. In Figure 4.8b, we present a cumulative histogram that we can use to infer probabilities of interest. For example, we can conclude that about 86% of the claims have about 40 or fewer days in accounts receivable.

Figure 4.8: Histograms created in Minitab 18 using the data in Table 4.3

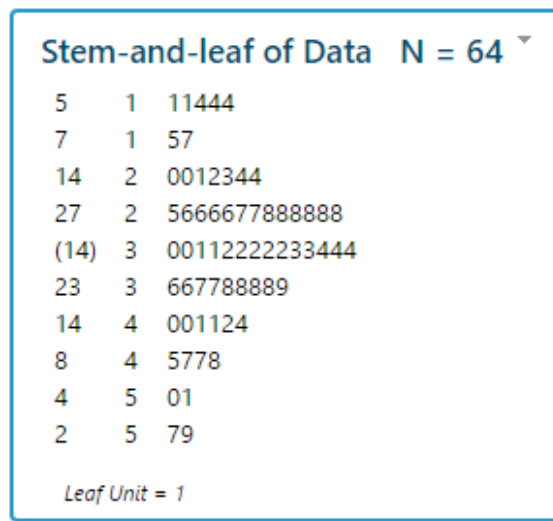
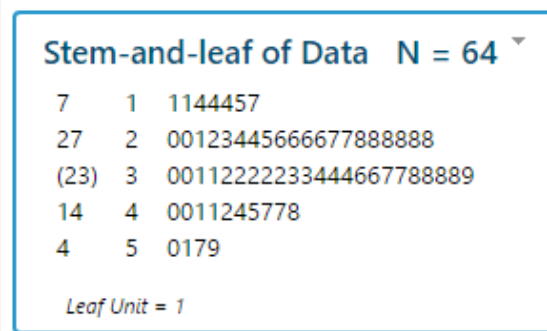
(a) A bell-shaped histogram



(b) A cumulative histogram



Using the same data, we created two stem-and-leaf diagrams, as portrayed in Figure 4.9.

Figure 4.9: Stem-and-leaf diagrams created in Minitab 18 using data in Table 4.3**(a)** A half-stem diagram (increment 5)**(b)** A full-stem diagram (increment 10)

The first column of the stem-and-leaf diagrams in Figure 4.9 portrays the cumulative count of leaves from both ends until we reach the median. Minitab uses parentheses () to show a leaf that contains the median. If the median results from two leaves, Minitab does not display the parentheses. Using Equation 4.24, we can find the position of the median, of the 50th percentile, as follows:

$$\begin{aligned}
 \text{position} &= (\text{percentile}) * (n + 1) \\
 &= 0.50 * (64 + 1) \\
 &= 32.5
 \end{aligned}$$

If we were interested in the position of the 90th percentile, we would find it like this:

$$\begin{aligned}
 \text{position} &= (\text{percentile}) * (n + 1) \\
 &= 0.90 * (64 + 1) \\
 &= 58.5
 \end{aligned}$$

We should note that the position of a percentile is different than the actual value of the percentile. For example, to determine the actual value of the 50th percentile in this example, we first have to find the difference between the values in the 32nd and 33rd positions and then multiply that difference with the decimal in the percentile position (0.5 in this example). We then add the result to the value in the 32nd position. In our example, the values in the 32nd and 33rd positions are 32 and 32, respectively. So, the value of the 50th percentile is $0.5 * (32 - 32) + 32 = 32$.

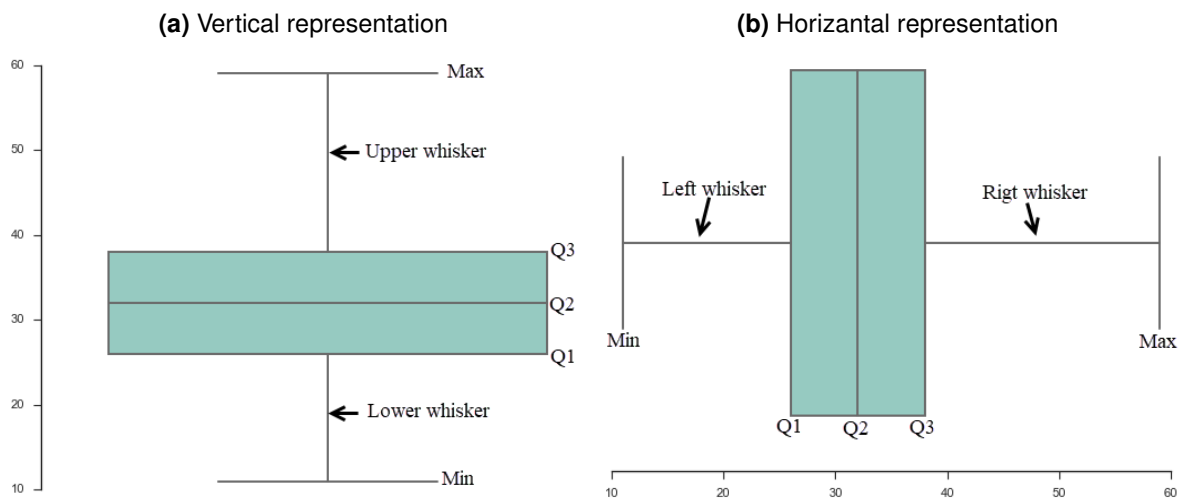
4.3.4 Box plots

We use **box plots**, also called **box and whisker** diagrams, to assess variability in the process in terms of the skewness, symmetry, and potential outliers in the data. We also can use box plots to estimate the central tendency of the process using the mean or median statistics.

Constructing a box plot

To construct a box plot, we connect a rectangular box in the middle of two opposing straight lines that we call whiskers. The outer end of a whisker represents the minimum or maximum value of the data. The inner ends that connect to the box coincide with the first quartile (Q1) or the third quartile (Q3). We also draw a perpendicular line inside the box to represent the second quartile (Q2) or the median of the data. Figure 4.10 portrays two typical representations of box plots.

Figure 4.10: Horizontal and vertical representations of box plots



We refer to the difference between Q3 and Q1 as the **interquartile range (IQR)**. We use the latter quantity to characterize the amount of variability in the data [43]. The bigger the IQR, the more variability in the process. The asymmetry of whiskers implies skewness toward the longer whisker. In some statistical software such as Minitab, the whiskers only have a length of $1.5 \times \text{IQR}$. We consider any data point that falls beyond this length, an outlier.

How-To 4.10 (Creating a box plot in Minitab 18)

To create a box plot in Minitab, click on *Graph > Boxplot > select the type of boxplot > select your data >> OK*.

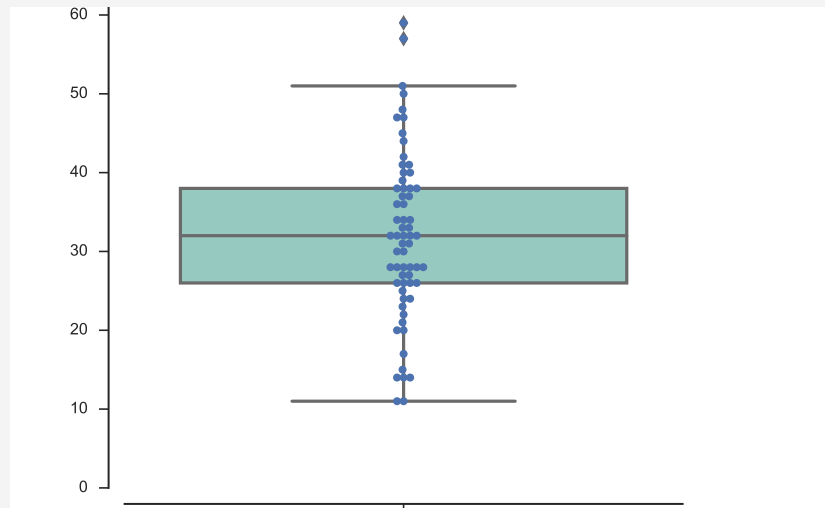
How-To 4.11 (Finding quartiles in Excel 2013)

To find quartiles in Excel 2013, use this function: `=QUARTILE.INC(yourdata, quartile_number)`

Example 4.8 (Box plot)

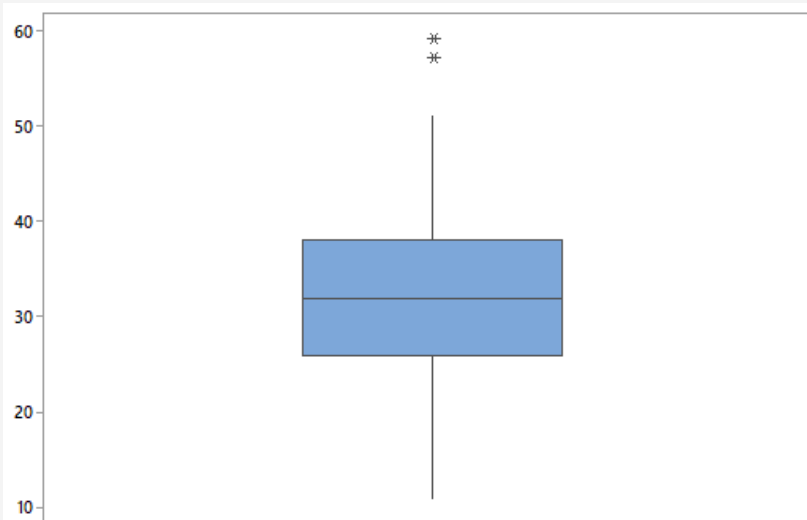
Using the data in Table 4.3, we drew a box plot that we portray in Figure 4.11.

Figure 4.11: A box plot drawn using the Python software and the data in Table 4.3



We determined the quartiles to be $Q1 = 26$, $Q2 = 32$, and $Q3 = 38$. This meant that $IQR = 38 - 26 = 12$. We drew the whiskers with a length of $1.5 \cdot IQR = 12 \cdot 1.5 = 18$. As a result, we appreciated two potential outliers, which suggest slight right skewness. Without these outliers, the data would be symmetric around the median. We used Seaborn, a module for the Python software, to draw the box plot in Figure 4.11, which allowed us not only to show outliers but also the *swam points* [58]. Figure 4.12 portrays a box plot drawn using Minitab, which also shows the outlier points. Excel does not have convenient options to draw box plots.

Figure 4.12: A box plot drawn using Minitab based on the data in Table 4.3



4.4 Process capability

During quality improvement efforts, we conduct **capability analysis** to help answer the question of whether our process is capable of meeting the customer's specification limits or standards. For example, in health care, we want to know whether our process can meet various guidelines from regulating and accrediting agencies. An important pre-requisite to capability analysis is the **stability** of the process.

Definition 4.5 Capability analysis is a technique that involves measuring the uniformity of the process, given the specification limits and variability in the process [43].

Several techniques for conducting capability analysis exist. Here, we emphasize the use of graphical techniques such as histograms and process capability ratios (PCRs) like C_p .

4.4.1 The C_p ratio

Given the customer's upper specification limit (USL) and lower specification limit (LSL), we utilize the C_p ratio to measure the **potential** capability of a process. We formulate this ratio as follows:

$$C_p = \frac{USL - LSL}{6\sigma} \quad (4.25)$$

where σ represents the process variability in standard deviation units. We obtain 6σ from the sum of 3σ below the mean μ and 3σ above the mean. We recall that 99.73% of the normally distributed process falls within $\mu \pm 3\sigma$. Since μ is not part of the formula in Equation 4.25, we cannot conclude anything about the process mean using the C_p ratio. To make such a conclusion, we apply other PCRs such as C_{pu} , C_{pl} , and C_{pk} .

4.4.2 The C_{pu} , C_{pl} , and C_{pk} ratios

Like the C_p ratio, the C_{pu} , C_{pl} , and C_{pk} ratios also allow us to measure the capability of the process. But unlike the C_p ratio, these new ratios account for the process mean μ .

1. The C_{pu} ratio, formulated as

$$C_{pu} = \frac{USL - \mu}{3\sigma}, \quad (4.26)$$

measures the process capability of satisfying the USL given the mean μ and 3σ .

2. The C_{pl} ratio, formulated as

$$C_{pl} = \frac{\mu - LSL}{3\sigma}, \quad (4.27)$$

measures the process capability of meeting the LSL given the mean μ and 3σ .

3. The C_{pk} ratio, computed as

$$C_{pk} = \min(C_{pu}, C_{pl}), \quad (4.28)$$

measures the **actual** capability of the process. Here, $\min()$ symbolizes the minimum function.

When $C_{pk} = C_p$, we deduce that the process mean μ is about centered between USL and LCL . When $C_{pk} < C_p$, the process is off-center. For cases of $C_{pk} < 0$, the process is completely outside of the specification limits.

4.4.3 Implementing PCRs

Before we can implement PCRs, we must be given the process standard deviation σ or be able to estimate this statistic from the process sample. We also may need to know the process mean μ . We can estimate both of these parameters, as previously discussed in Subsection 4.2.6.

Interpretation

In general, any **PCR greater than 1** implies that the process is capable of meeting the given standard. Otherwise, to improve the capability measure of interest, we must decrease the variability in the process given by σ . For an existing process, it is recommended that a PCR measure be at least 1.33 in a two-sided measure or 1.25 in a one-sided measure. For a new process, the recommended PCR values are 1.50 for a two-sided measure and 1.45 for a one-sided measure. In a two-sided, centered, and normally distributed process, a C_p of 0.25 implies 453,255 ppm², a C_p of 1 implies 2,700 ppm, and a C_p of 2 signifies about 0.0018 ppm [43]. Another practical interpretation of the C_p ratio follows from:

$$P = \left(\frac{1}{C_p} \right) \quad (4.29)$$

where P is the percentage of the specification bandwidth being used by the process.

1. If $P \leq 1$: the process is capable of meeting the specification limits.
2. If $P > 1$: the process is using more than the allowed bandwidth and is therefore not capable of meeting the specification limits.

Besides measuring the process capability, the C_p ratio can also be utilized to evaluate the acceptability of the *gauge* instrument, as we discuss further in Module 7.

Remark: When a process is **not** in control, performance indices like P_p , P_{pu} , P_{pl} , and P_{pk} are sometimes used to measure the capability. These indices are essentially similar to the PCRs



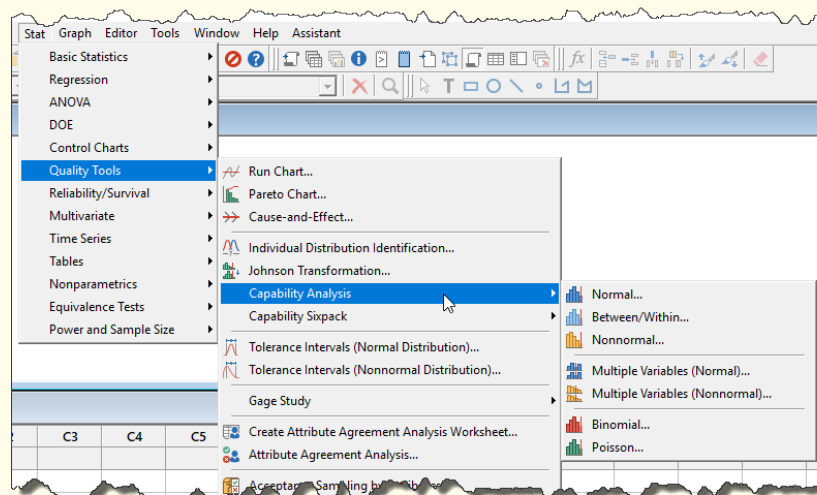
²ppm: defective parts per million

discussed earlier except that σ is replaced with the process standard deviation s . These ratios should be applied with caution, or not be used at all, due to possible misleading conclusions that may result when the process is not stable [43].

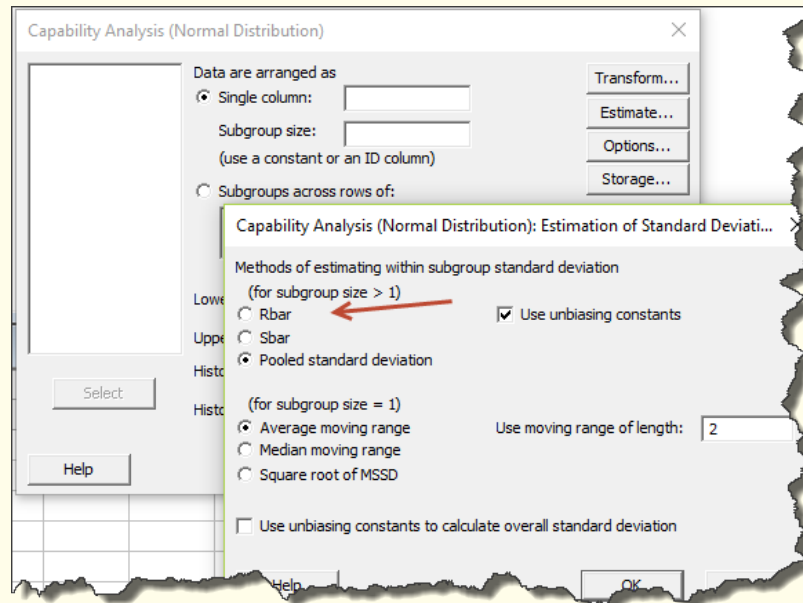


How-To 4.12 (PCRs in Minitab 18) Click on *Stat > Quality Tools > Capability Analysis > Normal > upload your data*. See the snapshot in Figure 4.14. You can change the distribution as necessary.

Figure 4.13: Capability analysis options in Minitab 18



For additional graphical results, click on *Stat > Quality Tools > Capability Sixpack > select the appropriate distribution > upload your data > OK*. To change the σ estimation method, click on *Estimate > Select the preferred technique* (see the snapshot in Figure 4.14)

Figure 4.14: σ estimation options in Minitab 18**Example 4.9 (Capability analysis)**

For the past 40 days, the medical coding manager at Metro City Hospital has been sampling five inpatient records daily (x_1, \dots, x_5) to track the processing times in minutes (mins). The manager would like to test the process capability to meet the rate of USL = 45 minutes and LSL = 15 minutes. The process is statistically stable. Table 4.4 shows the data collected so far.

Table 4.4: Processing times of inpatient records at Metro City Hospital in 2018

<i>Sample</i>	x_1	x_2	x_3	x_4	x_5	<i>Sample</i>	x_1	x_2	x_3	x_4	x_5
1	32	25	23	37	30	21	27	23	33	24	32
2	37	31	35	36	33	22	31	33	37	30	32
3	37	27	30	33	26	23	30	36	31	26	25
4	36	32	27	21	31	24	36	28	30	26	30
5	34	40	40	26	30	25	23	22	28	17	32
6	32	33	38	30	29	26	30	26	37	32	23
7	20	35	34	28	22	27	32	30	31	28	34
8	35	21	35	33	25	28	29	36	30	25	27
9	36	29	41	27	26	29	27	32	27	32	30
10	33	27	35	31	20	30	29	24	23	26	29
11	30	29	33	24	31	31	28	30	25	27	31
12	31	27	27	32	35	32	31	31	28	30	31
13	26	24	38	32	25	33	24	26	37	30	35
14	37	26	21	41	31	34	26	31	30	25	28
15	33	24	31	33	29	35	28	26	25	38	26
16	26	30	35	32	22	36	30	29	30	30	28
17	29	37	34	34	24	37	32	33	28	25	35
18	33	34	25	33	31	38	21	38	32	30	30
19	26	30	29	26	41	39	35	27	31	34	37
20	33	24	33	32	35	40	33	45	23	28	39

To help the manager analyze the capability of this process, we start by estimating parameters σ and μ . We approximate μ using $\bar{\bar{x}}$, which is the average of all sample averages. To approximate σ , we utilize Equation 4.15 since $n \leq 10$. Using Excel, we obtain the range by $MAX(sample) - MIN(sample)$. We calculate the average of each sample using the $AVERAGE(sample)$ function. We illustrate the setup of our spreadsheet in Figure 4.15.

Figure 4.15: A spreadsheet setup for calculating the ranges and averages of the data in Table 4.4

	A	B	C	D	E	F	G	H
1	<i>Sample</i>	x_1	x_2	x_3	x_4	x_5	Range	Mean
2	1	32	25	23	37	30	14	29.4
3	2	37	31	35	36	33	6	34.4
4	3	37	27	30	33	26	11	30.6
...
40	39	35	27	31	34	37	10	32.8
41	40	33	45	23	28	39	22	33.6
							11.15	30.115

About Figure 4.15, we obtained the range of sample 1 by $= MAX(B2 : F2) - MIN(B2 : F2) = 14$. We obtained the average of sample 1 by $= AVERAGE(B2 : F2) = 29.4$. We applied the same formulas to other samples. In the end, we calculated \bar{R} using $= AVERAGE(G2 : G41) = 11.15$ and $\bar{\bar{x}}$ using $= AVERAGE(H2 :$

$H41) = 30.115$. To approximate σ , we proceeded as follows:

$$\sigma \approx \hat{\sigma} = \frac{\bar{R}}{d_2} = \frac{11.15}{2.326} = 4.80 \quad (4.30)$$

We obtained the value of $d_2 = 2.326$ from Appendix Table 1 when $n = 5$. We proceeded to calculate the PCRs of interest as follows:

$$C_p = \frac{USL - LSL}{6\sigma} = \frac{45 - 15}{6(4.80)} = 1.043 \quad (4.31)$$

$$C_{pu} = \frac{USL - \mu}{3\sigma} = \frac{45 - 30.115}{3(4.80)} = 1.035 \quad (4.32)$$

$$C_{pl} = \frac{\mu - LSL}{3\sigma} = \frac{30.115 - 15}{3(4.80)} = 1.051 \quad (4.33)$$

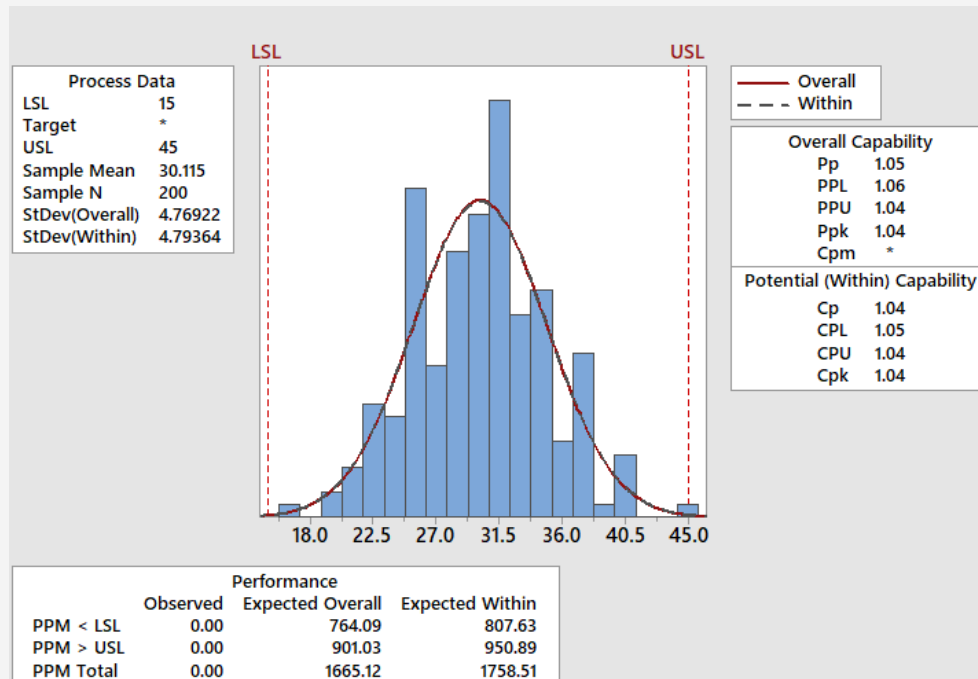
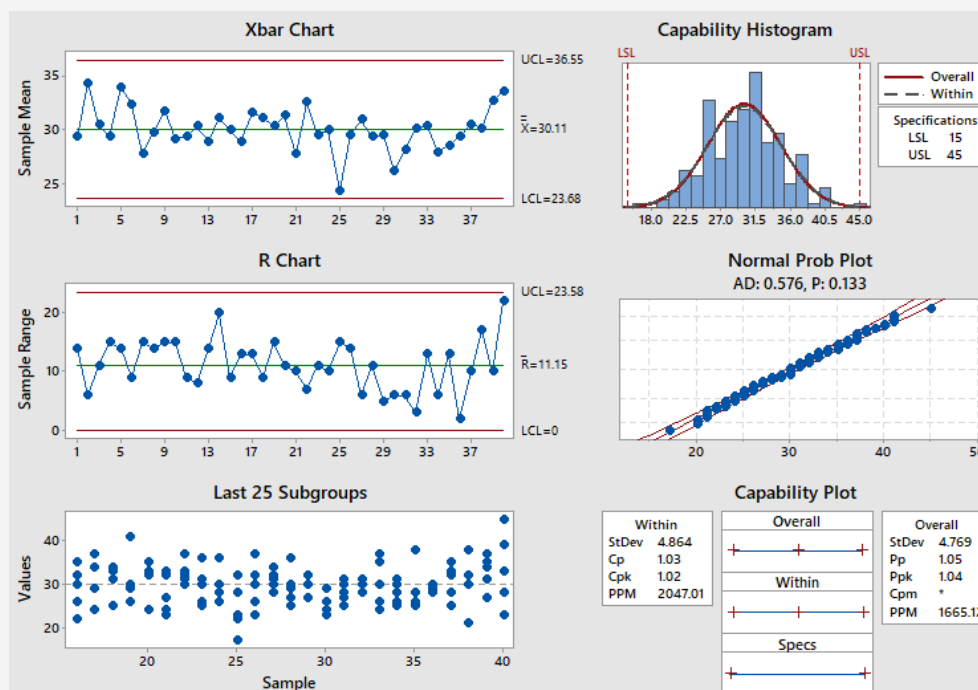
$$C_{pk} = \min(C_{pu}, C_{pl}) = \min(1.035, 1.051) = 1.035 \quad (4.34)$$

Given that all ratios are greater than 1, the process is capable. Furthermore, we can deduce that the process mean is about centered between USL and LSL, since $C_p \approx C_{pk}$. We estimate the amount of the bandwidth that the process is utilizing this way:

$$P = \left(\frac{1}{C_p} \right) 100 = \left(\frac{1}{1.043} \right) 100 = 96\%. \quad (4.35)$$

The 96% utilization rate of the allowed bandwidth is not desirable since it leaves little to no room for error. To improve the capability, the process variability must be decreased. Figure 4.16 portrays a Minitab capability report. The **histogram** in this figure confirms that the process is about centered, but is using almost all of the allowed bandwidth between LSL and USL. The **Capability Plot** substantiates this claim. The intervals in this plot portray different estimations of the spread in the process. The **within** interval follows from the unbiased estimation of σ using sample data.

Figure 4.17 depicts additional Minitab results related to the sixpack report. The **Xbar and R** charts in Figure 4.17 indicate that the process is stable since points move randomly around the centerline, and no point falls outside of the control limits. The **Last 25 Subgroups** chart validates the conclusion about the random behavior of the sample mean since no clear pattern is discernible. But, we do notice that sample 40 seems to show more variability than other points. The **Normal Probability Plot** portrays points falling alongside the fitted line, an indication that the process is about normally distributed.

Figure 4.16: Process capability report in Minitab 18 based on the data in Table 4.4**Figure 4.17:** Process capability sixpack report in Minitab 18 based on the data in Table 4.4

4.5 Check your understanding

1. What software would you recommend **to create histograms** for continuous data?
2. Justify the need for assessing the variability and capability of the process.
3. **What is the position** of the 50th percentile when we have 91 data points?
4. **What is the position** of the 90th percentile when we have 91 data points? If the values in the 82nd and 83rd positions are 31 and 56, **what is the actual value** of the 90th percentile?
5. In Minitab, what is the difference between an **increment of 10** and an **increment of 5** when creating a stem-and-leaf diagram?
6. You are comparing the process of the **release of information (ROI)** between two comparable hospitals. In hospital A, you found IQR of 50. In hospital B, you found IQR of 100. What can you conclude about the quality of ROI in these hospitals?
7. What is the significance of a **whisker** that is **$1.5 \times \text{IQR}$** long?
8. What rule of thumb can you apply to determine the **number of bins** in a histogram?
9. What is the consequence of having **too many or too few bins** in a histogram?
10. What conclusion can you generally draw if you had a histogram with **two modes**?
11. Which sample statistic that considers the **peakedness** and **tailedness** of the process?
 - (a) Skewness
 - (b) Kurtosis
 - (c) Mean
 - (d) Variance
12. We determine **IQR** in a box plot using:
 - (a) $Q2 - Q1$
 - (b) $Q3 - Q2$
 - (c) $Q4 - Q3$
 - (d) $Q4 - Q1$
13. What statistical analysis can you perform to validate whether your process can meet the specifications of the customer?
14. From 25 samples, each with size 10, you estimated that the hospital's EHR down-time average time was 30 minutes with the standard deviation of 2 minutes. Your process is assumed stable and normally distributed.

- (a) What is the **potential** capability of the system to meet the specifications of $LSL = 10$ minutes and $USL = 45$ minutes?
 - (b) What is the **actual** capability of the system to meet the specifications of $LSL = 10$ minutes and $USL = 45$ minutes?
 - (c) What is the percentage of the allowed bandwidth is the process using?
 - (d) What is the probability of observing a downtime period longer than 45 minutes?
15. Using Minitab, conduct the capability analysis of the data given Example 4.9, given $LSL = 20$ and $USL = 50$. Attach your Minitab results, including the interpretation of all relevant findings. What additional insights can you obtain from the capability sixpack report?
16. When a process is in control, we apply the following ratio to measure the actual capability.
- (a) C_p
 - (b) P_p
 - (c) C_{pk}
 - (d) P_{pk}
17. If C_p is greater than 1:
- (a) The process is capable
 - (b) The process is not capable
18. We use $\min(C_{pu}, C_{pl})$ to calculate:
- (a) C_p
 - (b) C_{pk}
 - (c) C_{pu}
 - (d) C_{pl}
19. Given the 6,210 ppm value obtained during the capability analysis in Minitab, what is the estimated Motorola sigma level of this process?
- (a) 1
 - (b) 2
 - (c) 3
 - (d) 4
 - (e) 5
 - (f) 6

20. From an audit of 30 orders for the diagnostic test Z, you found the average entry time in the CPOE system to be 7 minutes with $\bar{R} = 3$. Given LSL = 5 minutes and USL = 10 minutes, what is the capability of this process? What can you conclude about the central tendency of this process?
21. Name one sample statistic that we can employ to measure variability in the process?
22. The term *subgroup* has the same meaning as
- (a) sample
 - (b) number
 - (c) population
 - (d) variance
23. The manager of Metropolis Clinic has received several patient complaints about excessive waiting times. To investigate the matter, the manager performed a retrospective audit of 5 random visits each day for 25 days. From this scenario, the sample size is:
- (a) 1
 - (b) 5
 - (c) 12
 - (d) 60
24. Which of the following is an example of a **statistic** measure of a process:
- (a) a sample mean
 - (b) a sample size
 - (c) a sample number
 - (d) a random sample
25. In _____ sampling, each item in the process has an equal chance of being selected:
- (a) random
 - (b) systematic
 - (c) stratified
 - (d) cluster
26. We use this formula $\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ to calculate the sample:
- (a) mean
 - (b) variance

- (c) standard deviation
 - (d) median
27. Your population has a mean $\mu = 50$ and $\sigma = 10$. If you took a sample of size 7 and observed a mean of 55, what is your sampling error?
- (a) 5
 - (b) 7
 - (c) 10
 - (d) 40
28. Your population has a mean $\mu = 50$ and $\sigma = 10$. If you took a sample of size 7 and observed a mean of 55, what is your standard error?
- (a) 3.8
 - (b) 7.0
 - (c) 11.9
 - (d) 17.6
29. The ratio between the two sample variances is likely to follow which distribution?
- (a) normal distribution
 - (b) t-distribution
 - (c) F-distribution
 - (d) Chi-square distribution
30. When sampling your process, you should seek to — within subgroup differences
- (a) maximize
 - (b) minimize
31. Which of the following sample statistic is a biased estimator of the related population parameter?
- (a) Mean
 - (b) Variance
 - (c) Standard deviation
 - (d) Sum of Squares
32. Which of the following distribution has the mean and variance parameters that are equal?
- (a) Geometric
 - (b) Poisson

- (c) Normal
 - (d) Bernoulli
33. Which statistic do you need to create a confidence interval for your population mean?
- (a) Standard error
 - (b) Sampling error
 - (c) Regression error
 - (d) Statistic error
34. Why are we likely to assume normal distribution during quality improvement projects?
35. What graphical technique can we use to verify the normality assumption?

MODULE 5

Tools for evaluating relationships in data

In this module, we review graphical and statistical tools for evaluating relationships in the process data. The visual tools we discuss include scatter and marginal plots. The statistical measures we consider include the Pearson and Spearman correlation coefficients. We also discuss linear regression models and goodness-of-fit measures. For software, we use Minitab and Excel.

Key concepts and tools: Relationship in data; Correlation; Scatter plots; Marginal plots; Pearson coefficient; Spearman coefficient; Dependent variable; Independent variables; Linear regression model; Hypothesis testing; Residuals plots; R^2 ; $R^2_{adjusted}$; P-value

Major objectives

After studying this module, you will be able to:

1. Define key concepts and tools for assessing the relationship in data
2. Explain the concept of correlation analysis
3. Contrast scatter plots to marginal plots
4. Implement the Pearson correlation coefficient
5. Calculate the Spearman correlation coefficient
6. Justify the need for regression modeling
7. Fit a linear regression model to data
8. Interpret residual plots for the goodness-of-fit
9. Differentiate between R^2 and $R^2_{adjusted}$
10. Conduct hypothesis tests

5.1 Introduction

We dedicate this module to the review of tools for evaluating relationships in data. We start with a discussion about scatter and marginal plots. We demonstrate how to create these plots using Excel and Minitab. Next, we expand our discussion to include correlation measures such as the Pearson and Spearman coefficients. We end the module with an overview of regression techniques.

5.2 Visual tools

5.2.1 Scatter diagrams

A scatter diagram is one of the major tools of statistical process control (SPC) that we use in the *Analyze* phase of DMAIC to evaluate the correlation between two variables.

Constructing a scatter diagram

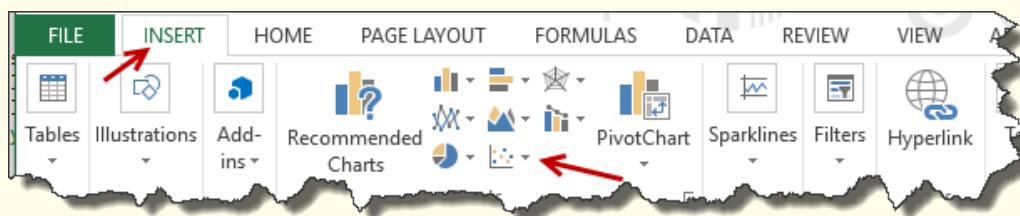
To create a scatter diagram, we start by organizing the data in the table format by storing each variable in a separate column. Each row constitutes an ordered pair of observations. To create a scatter plot, we simply graph a 2-dimensional chart of the variables in the Cartesian coordinates system.

How-To 5.1 (Creating a scatter plot in Excel 2013)

To create a scatter diagram in Excel 2013, click on *INSERT > Scatter Charts*. See the snapshot in Figure 5.1.



Figure 5.1: Options for a scatter plot in Excel 2013

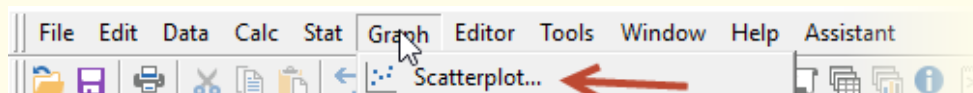


How-To 5.2 (Creating a scatter plot in Minitab 18)

To create a scatter diagram in Minitab 18, click on *Graph > Scatterplot*. See the snapshot in Figure 5.2).



Figure 5.2: Options for a scatter plot in Minitab 18



Interpreting a scatter diagram

We characterize the relationship in a scatter plot as follows:

1. **Positive correlation** (or direct relationship) when two variables increase together.
2. **Negative correlation** (or inverse relationship) when one variable increases as the other one decreases.
3. **No correlation** (or no relationship) when two variables do not vary together.

We can further differentiate the strength of the relationship by using terms such as **strong**, **moderate**, or **weak**. For example, we can describe two variables as being *strongly and positively correlated*. It is also common to fit a line on the scatter data to evaluate the **linearity** or **non-linearity** of the relationship. Figure 5.3 portrays examples of scatter plots. In Subfigure 5.3a, strong positive and linear relationship exists between variables X and Y. This association implies that as X increases, Y tends to increase too. In Subfigure 5.3b, we observe a strong negative and linear relationship between X and Y, implying that as X increases, Y tends to decrease. In Subfigure 5.3c, no obvious relationship is discernible, meaning that X and Y are likely independent variables. Subfigure 5.3a shows a non-linear relationship between X and Y, which, after fitting a line, seems quadratic.



Remark: Scatter diagrams do not suggest that one variable causes another, rather imply that a **potential relationship** exists. Other statistical tools, such as the design of experiments (DOE), are best suited for entertaining the **causality** issue [43].

5.2.2 Marginal plots

Marginal plots are scatter diagrams that, in addition to portraying the relationship in the data, depict the distributions of the variables under study. Minitab 18 provides options for displaying marginal plots using box plots, histograms, or **dot plots**. Dot plots are similar to stem-and-leaf diagrams, except that in the former leaves are made of dots instead of numbers.



How-To 5.3 (Creating a marginal plot in Minitab 18)

To create a marginal plot in Minitab 18, click on *Graph > Marginal Plot*. See the screenshot in Figure 5.4.

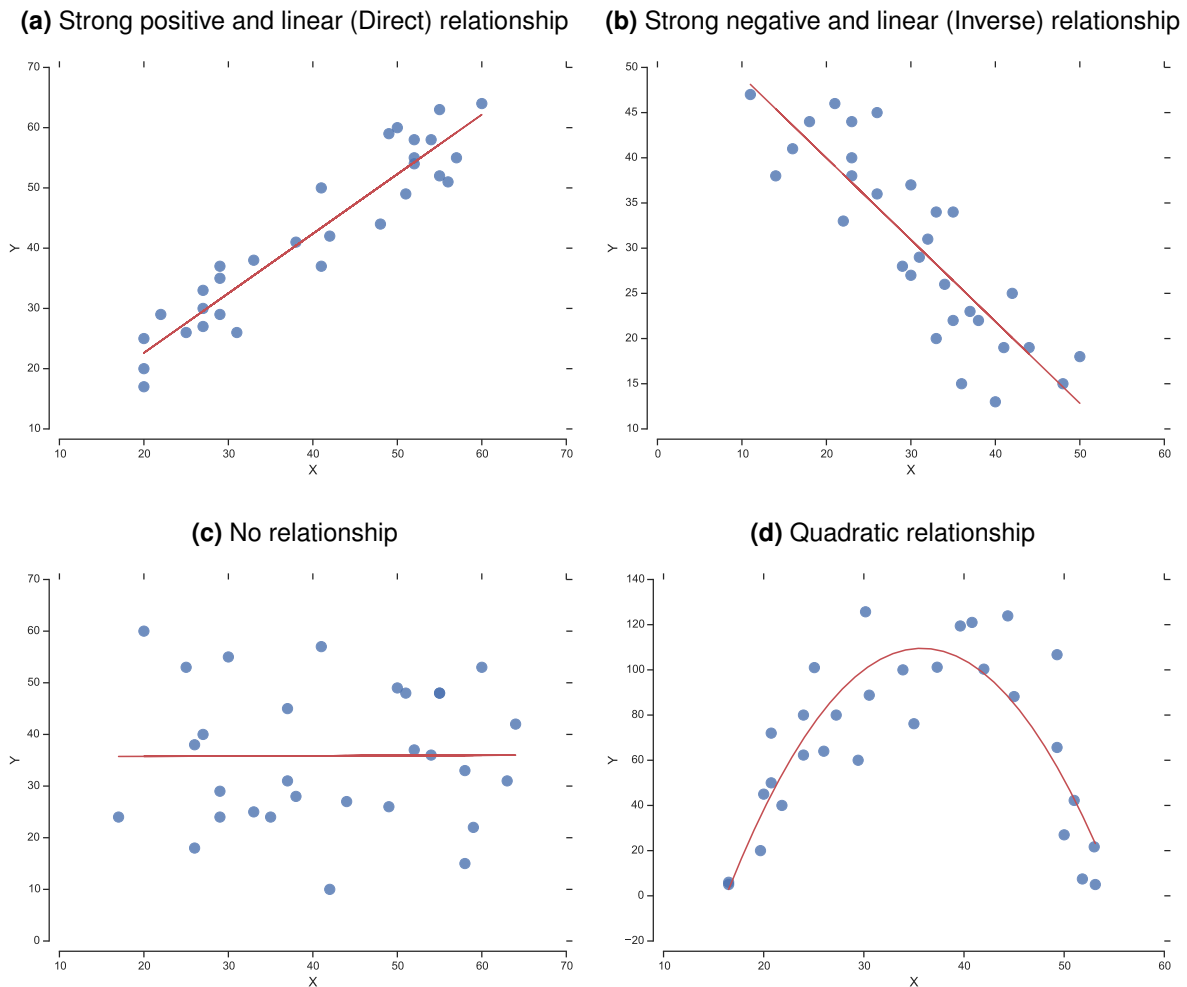
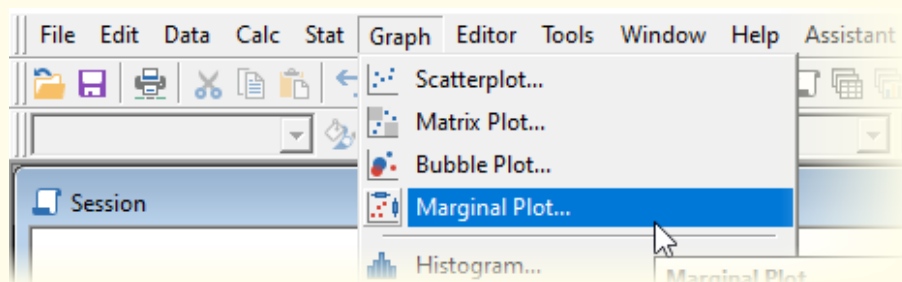
Figure 5.3: Examples of scatter diagrams**Figure 5.4:** Options for a marginal plot in Minitab 18**Example 5.1 (Marginal plots)**

Table 5.1 shows audit results from an informatics process at Metropolis Hospital. Using these data and Minitab 18, we created a marginal plot with a box-plot as portrayed in Figure 5.5a. Figure 5.5b shows a marginal plot of histograms. In both

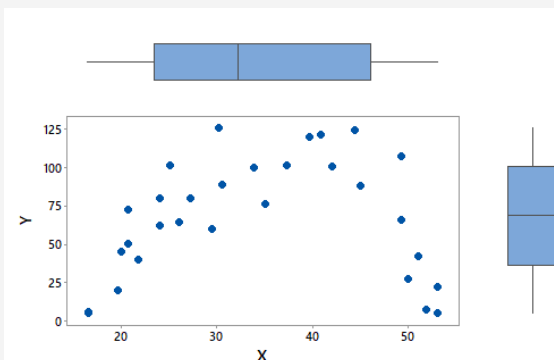
of these plots, we appreciate that the marginal distribution of X is somewhat right-skewed, whereas the marginal distribution of Y is about symmetric with slight lower skewness.

Table 5.1: Audit results from an informatics process at Metropolis Hospital

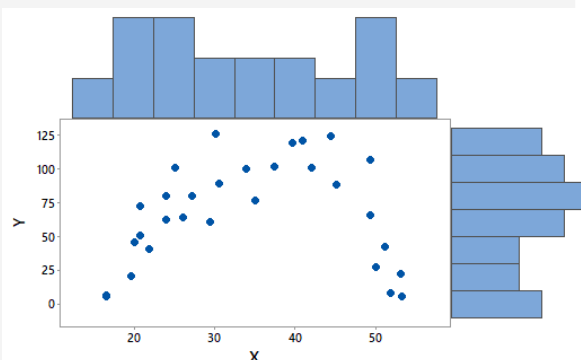
<i>Sample</i>	<i>X</i>	<i>Y</i>	<i>Sample</i>	<i>X</i>	<i>Y</i>
1	53.11	5	16	26	64
2	51.82	7.46	17	33.9	100
3	53	21.67	18	25.06	101
4	51	42.22	19	29.43	60
5	49.29	65.66	20	50	26.99
6	45	88.18	21	23.97	80
7	49.29	106.72	22	21.82	40
8	39.63	119.43	23	27.24	80
9	30.17	125.7	24	23.97	62.3
10	44.37	123.89	25	20.75	50
11	40.8	121	26	16.5	6
12	41.98	100.32	27	20.75	72
13	37.32	101.18	28	20	45
14	30.54	88.8	29	16.5	5
15	35	76.16	30	19.68	20

Figure 5.5: Marginal plots created using Minitab 18 and the data in Table 5.1

(a) A box plot marginal plot



(b) A histogram marginal plot



5.3 Statistical measures

Besides visual tools, we can also use statistical measures to establish the degree of the relationship between two variables. Common statistical measures for this purpose are the Pearson and Spearman correlation coefficients. In addition, we can use regression

models to evaluate the relationship between the dependent and independent variables in the process.

5.3.1 Correlation analysis

We conduct a correlation analysis to evaluate the relationship between two variables. Here, we show how to achieve this goal using the Pearson and Spearman correlation coefficients. Another similar technique that we do not cover in this section is the Kendall rank correlation coefficient [2].

Pearson correlation coefficient

The Pearson correlation coefficient is a parametric technique that we use to quantify the degree of the relationship between two continuous variables. We calculate this coefficient by taking the ratio between the corresponding covariance and standard deviations statistics. To demonstrate, let's assume that we have two samples X and Y , each with size n , and means of \bar{X} and \bar{Y} , respectively. The covariance of X and Y , $cov(X, Y)$ is given by:

$$cov(X, Y) = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{n - 1} \quad (5.1)$$

where $n - 1$ represents the degrees of freedom. We obtain the Pearson correlation coefficient $\rho(X, Y)$, sometimes denoted as r , this way:

$$\rho(X, Y) \equiv r = \frac{cov(X, Y)}{s_X s_Y} \quad (5.2)$$

Here, s_X is the standard deviation of variable X and s_Y is the standard deviation of variable Y . It follows that [37]:

$$-1 \leq r \leq 1 \quad (5.3)$$

The more r is close to -1 , the more X and Y are **negatively correlated**. The more r is close to 1 , the more X and Y are **positively correlated**. When r is around 0 , the two variables have little to no correlation.

Spearman coefficient

The Spearman correlation coefficient is a non-parametric technique that uses ranks instead of the actual values of the variables under study. Given two variables X and Y , we replace a given value of x in X by its rank $R(x)$ and a given value of y in Y by its rank $R(y)$. We denote the Spearman coefficient as r_S and compute it this way:

$$r_S = \frac{cov(R(X), R(Y))}{s_{R(X)} s_{R(Y)}} \quad (5.4)$$

where $-1 \leq r_S \leq 1$. In the simplified case of no ties in ranks, we compute r_S as follows:

$$r_S = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (5.5)$$

where $d_i = R(x_i) - R(y_i)$ [57]. We typically use the Spearman coefficient to assess the degree of a monotonic relationship between two variables. Just like in Pearson's case, the more r_S is close to -1 , the more X and Y are **negatively correlated**. The more r_S is close to 1 , the more X and Y are **positively correlated**. When r_S is around 0 , the two variables are likely not to be correlated.



How-To 5.4 (Covariance and Correlation in Excel 2013) To calculate $cov(X, Y)$ in Excel, use this function `=COVARIANCE.S(X, Y)`. To calculate $\rho(X, Y)$, use this function `=CORREL(X, Y)`.



How-To 5.5 (Covariance and Correlation in Minitab 18) In Minitab, you can calculate the covariance and correlation between two variables by clicking on *Stat > Basic Statistics > chose Correlation... or Covariance... > select your data > OK*.

5.3.2 Regression

We use regression techniques to create a mathematical model of the relationship between dependent and independent variables. Here, we only consider a **linear regression** model with one dependent variable.

One dependent variable and one independent variable

Let's assume that we have two variables X and Y , where Y is the dependent variable and X the independent variable. The corresponding linear regression model follows.

$$Y = \beta_0 + \beta_1 X + \epsilon \quad (5.6)$$

Both β_0 and β_1 are regression coefficients. The ϵ term denotes the regression error that we refer to as residuals. We estimate β_1 using $\hat{\beta}_1$, as follows:

$$\hat{\beta}_1 = r \frac{s_Y}{s_X} \equiv \frac{cov(X, Y)}{var(X)} \quad (5.7)$$

where $var(X)$ is the variance of X and r is the Pearson correlation coefficient between X and Y . We calculate $\hat{\beta}_0$ to approximate β_0 this way:

$$\hat{\beta}_0 = \bar{Y} - \beta_1 \bar{X} \quad (5.8)$$

where \bar{Y} is the mean of Y and \bar{X} is the mean of X [37]. We can predict the values of Y using \hat{Y} , as follows:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X \quad (5.9)$$

The coefficient $\hat{\beta}_1$ represents the **expected change in \hat{Y} per unit change in X** . We can think of $\hat{\beta}_0$ as the intercept of the line \hat{Y} . We obtain the error term this way:

$$\epsilon = Y - \hat{Y} \quad (5.10)$$

The sum of squares of the error SS_{error} is given by ϵ^2 as follows:

$$\epsilon^2 = (Y - \hat{Y})^2 \quad (5.11)$$

$$\equiv SS_{error} \quad (5.12)$$

We assume ϵ to be a normally distributed random vector with mean zero and variance σ^2 . The unbiased estimator of σ^2 is given by $\hat{\sigma}^2$ and is calculated as follows:

$$\hat{\sigma}^2 = \frac{SS_{error}}{n - p} \quad (5.13)$$

where $n - p$ represents the degrees of freedom. Here, n is the size of Y and p is the number of independent variables k plus one for the intercept. That is, $p = k + 1$ [43].

One dependent variable and multiple independent variables

When we have more than one independent variable, say k , we create a multiple linear regression model as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \epsilon \quad (5.14)$$

where X_1, X_2, \dots, X_k are independent variables, $\beta_0, \beta_1, \dots, \beta_k$ are regression coefficients, and ϵ is the error term. From the least-squares method, we obtain the β coefficients as follows:

$$\hat{\beta} = (X'X)^{-1} X'Y \quad (5.15)$$

where $'$ symbol indicates transpose. Here, X is a matrix of size $n \times p$, Y is a vector of size $n \times 1$, and β is a vector of size $p \times 1$. Our \hat{Y} model looks as follows:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \cdots + \hat{\beta}_k X_k \quad (5.16)$$

The coefficient $\hat{\beta}_j$ (for $j : 0, 1, \dots, k$) represents the **expected change in \hat{Y} per unit change in X_j while holding all other independent variables constant**. If all independent variables were zero, \hat{Y} would equal β_0 . As before, $\epsilon = Y - \hat{Y}$ and $SS_{error} = \epsilon^2$

5.3.3 Goodness-of-fit

The techniques that we utilize to assess the goodness-of-fit of a linear regression model are the coefficient of determination, residual plots, and hypothesis testing.

Coefficient of determination

We denote the coefficient of determination using R^2 to represent the percent of the variation in Y that is explained by the variation in independent variables. We recall that:

$$SS_{error} = (Y - \hat{Y})^2 \quad (5.17)$$

Additionally, it follows that [36]:

$$SS_{regression} = (\hat{Y} - \bar{Y})^2 \quad (5.18)$$

$$SS_{total} = (Y - \bar{Y})^2 \quad (5.19)$$

We calculate R^2 as follows:

$$R^2 = \frac{SS_{regression}}{SS_{total}} \equiv 1 - \frac{SS_{error}}{SS_{total}} \quad (5.20)$$

where $0 \leq R^2 \leq 1$. The higher the value of R^2 , the better the goodness-of-fit of our model.



Remark: If we added more variables, necessary or unnecessary, to our regression model, SS_{error} would decrease per Equation 5.13, which would in turn increase R^2 per Equation 5.20. Accordingly, we must be careful when interpreting R^2 . We recommend using $R^2_{adjusted}$, instead of R^2 , to determine the goodness-of-fit of a regression model. We find $R^2_{adjusted}$ this way:

$$R^2_{adjusted} = 1 - \frac{SS_{error}/(n - p)}{SS_{total}/(n - 1)} \quad (5.21)$$

where $n - 1$ is the degrees of freedom of Y and $n - p$ is the degrees of freedom of the residuals [43].

Residual plots

Besides R^2 and $R^2_{adjusted}$, we can also use residual plots to assess the goodness-of-fit of our linear regression model. If a model is a good fit, we expect the residuals to be normally distributed with mean zero and a constant variance. So, the normal plot of the residuals should indicate that the points fall alongside the fitted line. Furthermore, we expect to see random behavior around a horizontal line when we plot ϵ versus \hat{Y} . Any non-random behavior indicates poor fit, which suggests non-linear behaviors [37]. Statistical packages such as Minitab also include p-values to facilitate the interpretation of the residual plots.

Hypothesis testing of regression models

In addition to the coefficient of determination and residual plots, we can also employ hypothesis testing to assess the goodness-of-fit of our regression model. In this test, we

conjecture that the partial regression coefficients of the model are zero, as follows:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0 \quad (5.22)$$

$$H_1 : \text{At least one } \beta_j \neq 0, \text{ for } j:1,2,\dots,k \quad (5.23)$$

Using the analysis of variance (ANOVA) technique, we obtain the **test statistic** as follows:

$$F_0 = \frac{MS_{regression}}{MS_{error}} \quad (5.24)$$

The **critical value** is given by:

$$F_{\alpha, k, n-k-1} \quad (5.25)$$

where α is the significance level, n is the size of Y , and k is the number of independent variables. A good-fit model will have at least one non-zero coefficient. That is, the test statistic will be greater than the critical value or the p-value less α . We could also test the hypothesis that the coefficient β_j (for $j : 1, \dots, k$) is zero using the t-test, as follows [36]:

$$H_0 : \beta_j = 0 \quad (5.26)$$

$$H_1 : \beta_j \neq 0 \quad (5.27)$$

We obtain the test statistic this way [36]:

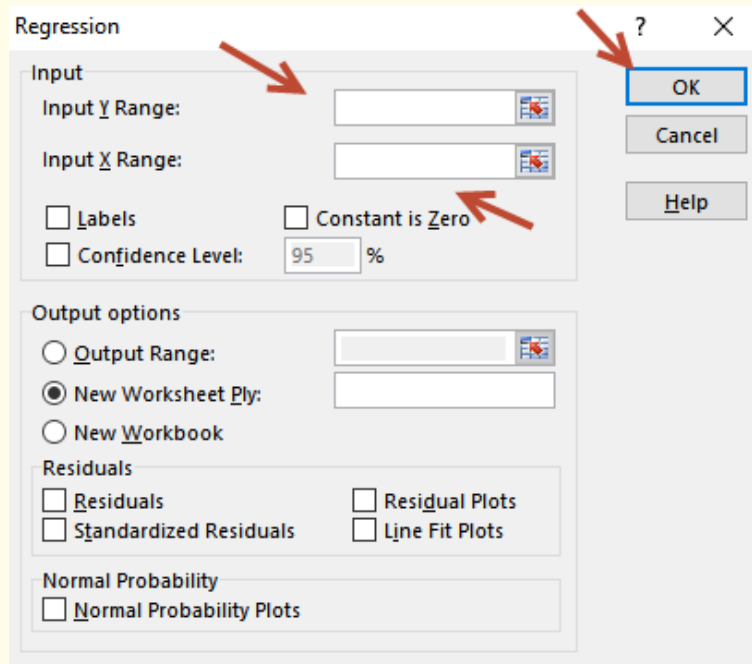
$$t = \frac{\beta_j - 0}{S_{\beta_j}} \quad (5.28)$$

where the *zero* in the Equation 5.28 comes from the null hypothesis and S_{β_i} is the standard error of the coefficient β_j . The degrees of freedom for this test are given by $n - p$, where n is the number of observations and $p = k + 1$, as before. We obtain the p-value for this test, as illustrated in How-To 5.7.

How-To 5.6 (Linear Regression in Excel 2013)

Open the Data Analysis add-in > *Regression* > *Input Y Range* for dependent variable > *Input X Range* for independent variables > Check *Labels* if you want to include the column headings > Check the boxes of *Residuals plots* to check for the model goodness-of-fit. See the snapshot in Figure 5.6.

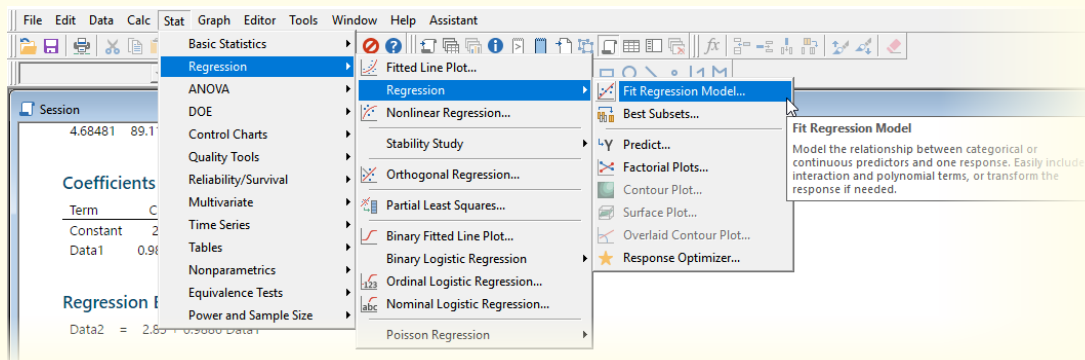


Figure 5.6: Regression using the Data Analysis add-in, Excel 2013**How-To 5.7 (The p-value of the t-test in Excel 2013)**

To obtain the p-value of the t-test in Excel 2013, use this equation: $=T.DIST.2T(ABS(t), n - p)$

**How-To 5.8 (Linear Regression in Minitab 18)**

Click on *Stat > Regression > Regression > Fit Regression Model...* On the screen that pops up, *input your Y data in the Responses and your X data in the Continuous predictors* > On the same screen, click on *Graphs > Check the boxes for Normal probability plot of residuals and Residuals versus fits* > *OK > OK*. See the snapshot in Figure 5.7.

**Figure 5.7:** Options for fitting a regression model, Minitab 18

Example 5.2 (Multiple linear regression) Table 5.2 presents data from a health informatics process at Central City Hospital, where Y is the dependent variable, and $X1$ and $X2$ are the independent variables. We fit a multiple linear regression model using Excel, and the results are displayed in Table 5.3. From these results, we construct the following model:

$$Y = 112.1 + 1.152X1 + 2.324X2 \quad (5.29)$$

We can interpret this model by saying that for each unit increase in $X1$, Y increases by 1.152. Likewise, for each unit increase in $X2$, Y increases by 2.324. Given that $R^2_{adjusted} = 87.80\%$, we can conclude that this model is a good fit. Additionally, given that the p-values of all coefficients are less than 0.05, we can infer that, at the 95% confidence level, all the model coefficients are non-zero. We should note that the p-values of the coefficients were obtained per How-To 5.7. For example, the p-value for $X1$ was obtained by $= T.DIST.2T(2.578985, 17) = 0.019508977$. All plots in Figure 5.8 suggest a good fit model, since residuals seem random and normally distributed. In the *Normal probability plot*, points fall along the fitted line, indicating the normality of the residuals. The *Histogram* plot also suggests the normal distribution, although this plot is not as informative since we need a lot of points to observe the bell shape, and here we only have 20 observations. The *Residuals Versus Fits* plot exhibits randomly scattered points, which is what is expected in a good fit model. The *Residuals Versus Order* plot also shows random behaviors since no particular pattern or trend is noticeable in the order of points.

Table 5.2: Data from a health informatics process at Central City Hospital

Y	X1	X2
87	1	11
293	72	41
394	116	59
277	71	39
185	28	19
353	50	51
151	30	19
268	61	36
296	82	40
262	56	30
195	49	30
471	136	81
239	43	29
481	72	87
377	101	57
313	93	34
203	49	21
481	145	129
490	139	89
218	36	21

Table 5.3: Excel Regression results for the data in Table 5.2

SUMMARY OUTPUT

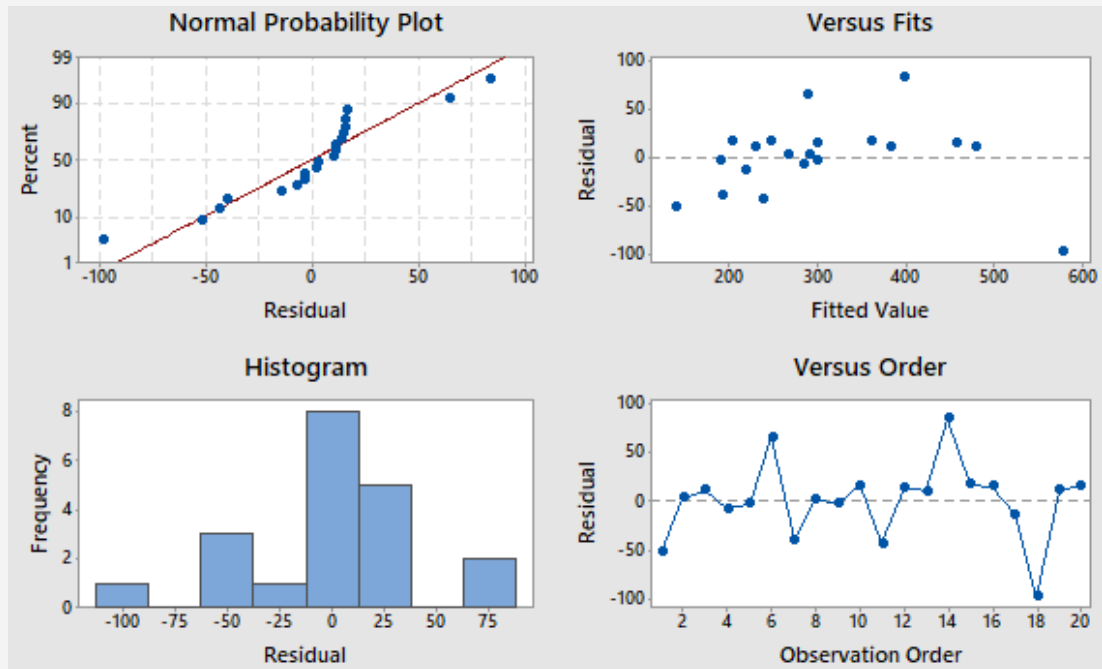
Regression Statistics

Multiple R	0.943824
R Square	0.890803
Adjusted R Square	0.877957
Standard Error	41.12373
Observations	20

ANOVA

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	2	234534.5	117267.2	69.34127	6.68E-09
Residual	17	28749.73	1691.161		
Total	19	263284.2			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
Intercept	112.0812	19.30276	5.806487	2.1E-05	71.35597	152.8065	71.35597	152.8065
X1	1.152057	0.446709	2.578985	0.019509	0.209582	2.094531	0.209582	2.094531
X2	2.323872	0.592855	3.919801	0.001103	1.073058	3.574686	1.073058	3.574686

Figure 5.8: Minitab graphical results for the regression of the data in Table 5.2

5.4 Check your understanding

1. What is the primary difference between marginal and scatter plots?
2. How do the Pearson and Spearman correlation coefficients differ?
3. From performing the correlation analysis between X and Y , you found $\rho = 0.6$. What can you conclude about the relationship between these variables?
4. Consider these two datasets: $A = \{7, 30, 25, 18, 13, 28, 1\}$ and $B = \{28, 3, 4, 15, 26, 25, 40\}$. What can you conclude regarding the relationship between A and B ?
5. You have taken two samples from your process. Suppose that you have found the sample covariance of 10 and the standard deviation of 10 in each sample. What can you conclude concerning correlation in your samples?
6. Which chart option in Excel is applicable when constructing histograms?
 - (a) Bar chart
 - (b) Pie chart
 - (c) Scatter plot
 - (d) Radar chart
7. What analytical tool is best suited for establishing causality?
 - (a) Scatter Plots
 - (b) Control charts
 - (c) DOE
 - (d) Regression models
8. What visual tool can you use to evaluate the potential relationship between two variables?
 - (a) Scatter plot
 - (b) Cause-and-effect diagram
 - (c) Box plot
 - (d) Stem-and-leaf diagram
9. How can you state the null hypothesis about the coefficients of a multiple regression model?
10. Why do we have to adjust the coefficient of determination?
11. Given the response Y and the fitted response \hat{Y} , how would you obtain the sum of squares of the residuals of your model?

12. How does the coefficient of correlation differ from the coefficient of determination?
13. Given the dependent variable Y and the independent variable X , how would you obtain the intercept of the corresponding linear regression model?
14. Given the residual plot from a regression model, how would you determine the goodness-of-fit?
15. Consider the following multiple linear regression model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \epsilon$. How can you interpret β_2 concerning Y ?
16. You have created a regression model from your sample data. How can you use the total sum of squares and the error sum of squares to obtain the coefficient of determination?
17. From fitting a multiple regression model in Excel, with 20 observations, you obtained the results in the following table:

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>
Intercept	89.75368	7.970321	11.26099
X1	1.7707	0.318356	5.562016
X2	1.669831	0.556528	3.000446

- (a) Write down the regression equation for this case
 - (b) What can you conclude about the significance of the coefficient of your regression model? Justify your answer using p-values
18. Consider the following dataset in Table 5.4.

Table 5.4

X	20	55	27	41	29	29	27	25	55	33	29	51	52	27	31	48	38	60	22	20	52	49	52	41	50	57	56	20	42	54
Y	17	63	27	37	37	35	33	26	52	38	29	49	58	30	26	44	41	64	29	25	55	59	54	50	60	55	51	20	42	58

- (a) Input the data in Excel or Minitab
 - (b) Fit a linear regression model
 - (c) Generate the plots of the residuals
 - (d) What can you conclude about your model in terms of:
 - i. R^2 or $R^2_{adjusted}$
 - ii. P-value
 - iii. The coefficients of the model
 - iv. Residual plots
19. Consider the following linear model:
- $$Y = 100 + 9X - 3Z \quad (5.30)$$

- (a) How would you interpret the coefficients of this model?
- (b) What information do you need to calculate the p-values of this model?
20. Fit a linear regression model to the data in Table 5.5, where Y is the dependent variable, and the rest of the variables are independent. What can you conclude from your results?

Table 5.5: Data from a health informatics process

Y	X1	X2	X3
250	35	3	0
360	29	4	1
165	36	7	0
43	60	6	0
92	65	5	0
200	30	5	0
355	10	6	1
290	7	10	1
230	21	9	0
120	55	2	0
73	54	12	0
205	48	5	1
400	20	5	1
320	39	4	1
72	60	8	0
272	20	5	1
94	58	7	0
190	40	8	1
235	27	9	0
139	30	7	0

MODULE 6

Tools for root cause and risk analysis

In this module, we review tools for evaluating the causes and risks of quality problems in the process. The tools and techniques that we cover include cause-and-effect diagrams, NGT¹ Q-sort, force-field analysis, FMEA², Is-Is Not matrix, and Why diagrams. For software, we use Minitab, Excel, and Visio.

Key concepts and tools: Brainstorming; Affinity grouping; Root cause analysis; Risk analysis; Cause-and-effect diagrams; Crowdsourcing; Force-field analysis; Voting techniques; Prioritization; NGT; FMEA; Why diagrams; Is-Is Not matrix

Major objectives

After studying this module, you will be able to:

- 1. Define key concepts and tools for cause and risk analysis*
- 2. Justify the need for root cause analysis*
- 3. Justify the need for risk analysis*
- 4. Implement brainstorming techniques*
- 5. Implement cause-and-effect diagrams*
- 6. Implement risk analysis techniques*
- 7. Apply affinity grouping techniques*
- 8. Implement a force-field analysis*
- 9. Survey prioritization techniques for causes and risks of quality problems*
- 10. Articulate examples of risks of quality problems in health care*

¹**NGT:** Nominal Group Technique

²**FMEA:** Failure Modes and Effects Analysis

6.1 Introduction

Root cause and risk analysis are major activities of quality improvement that we typically implement during the *Analyze* phase of DMAIC to attempt to identify potential causes and risks of quality problems.

6.2 A general procedure

The approach that one takes to conduct root cause and risk analysis varies by the situation, but the following procedure applies in typical processes:

1. Generate ideas about the causes and risks of the effect
2. Organize ideas into classes or themes using graphical or tabular tools
3. Rank and prioritize ideas

6.3 Generating ideas

We can generate ideas about the root causes and risks of quality problems by talking to the stakeholders of the process or by involving experts. Traditional techniques for this purpose include:

- **Brainstorming** sessions where we create a team of stakeholders to generate ideas about the likely causes and possible solutions to a given quality problem [49].
- **Delphi surveys**, where we involve a panel of experts and request their feedback about possible root causes and feasible solutions to a particular quality issue [43].

A more recent technique for generating ideas about root causes, risks, and potential solutions to quality problems is **crowdsourcing**, where we involve the general public via the web. In many cases, respondents participate for free and may not possess traditional credentials in the fields of interest. Yet, this new way of generating ideas has proven to be very effective [40].

6.4 Organizing ideas

In this section, we discuss techniques for organizing ideas about root causes, risks, and potential solutions to a quality problem.

6.4.1 Affinity grouping

Affinity grouping is an organizational technique for classifying similar ideas into major themes [49]. The same team that proposed original ideas could help with the classification task. For the case of a large number of suggestions, we could initially utilize *machine learning* algorithms to conduct **sentiment analysis** using embedding methods and other topic modeling techniques such as the Latent Dirichlet Allocation method [9].

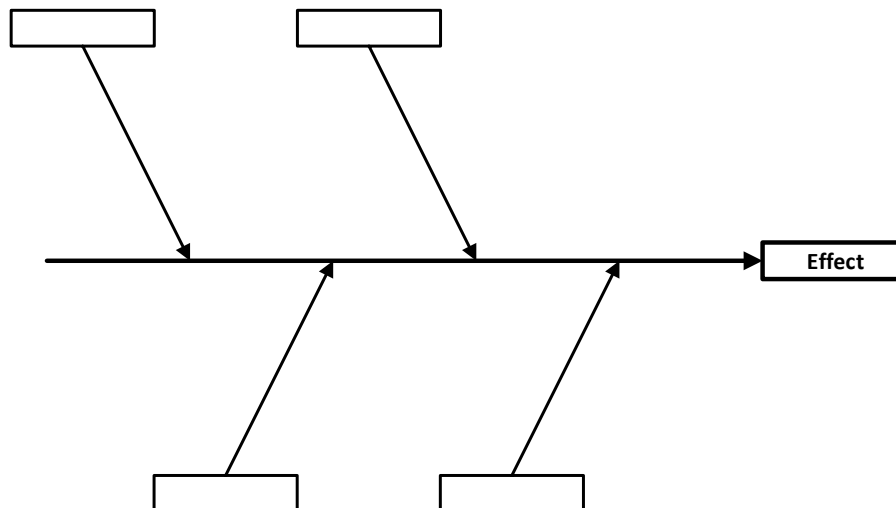
6.4.2 Cause-and-effect diagrams

A **cause-and-effect diagram** is one of the primary tools of SPC³ that we utilize to organize potential causes by the major process inputs such as *people, policy, methods, and materials* [50]. Other names of this diagram are **Ishikawa** (the name of an individual who first introduced it [43]), and **fishbone** (since the diagram looks like the skeleton of a fish) [43]. A basic algorithm for creating a cause-and-effect diagram follows.

Step 1: Draw a horizontal arrow from left to right and connect it to a box of the effect. This line will serve as the main branch.

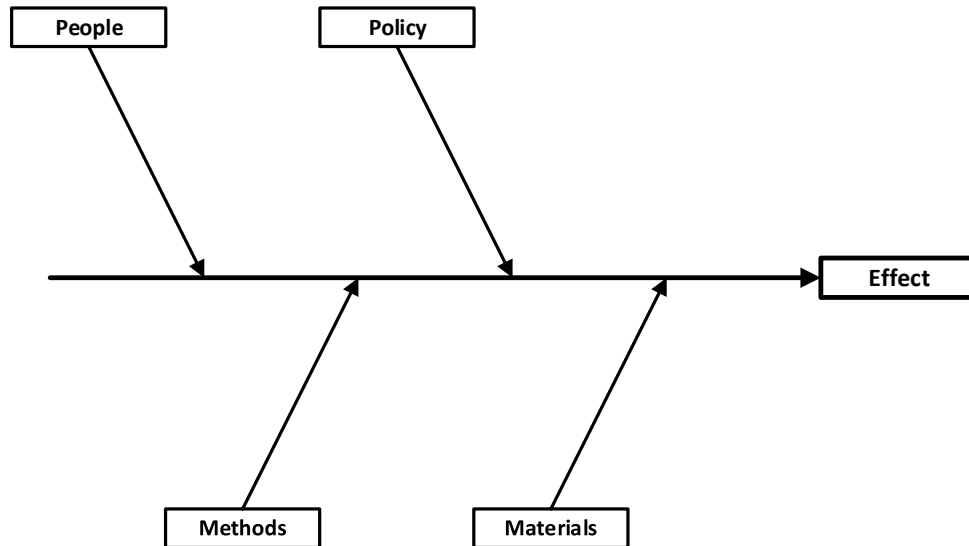


Step 2: Draw boxes for the process key inputs and connect them to the main branch using right-slanted arrows. We refer to these secondary arrows as sub-branches.

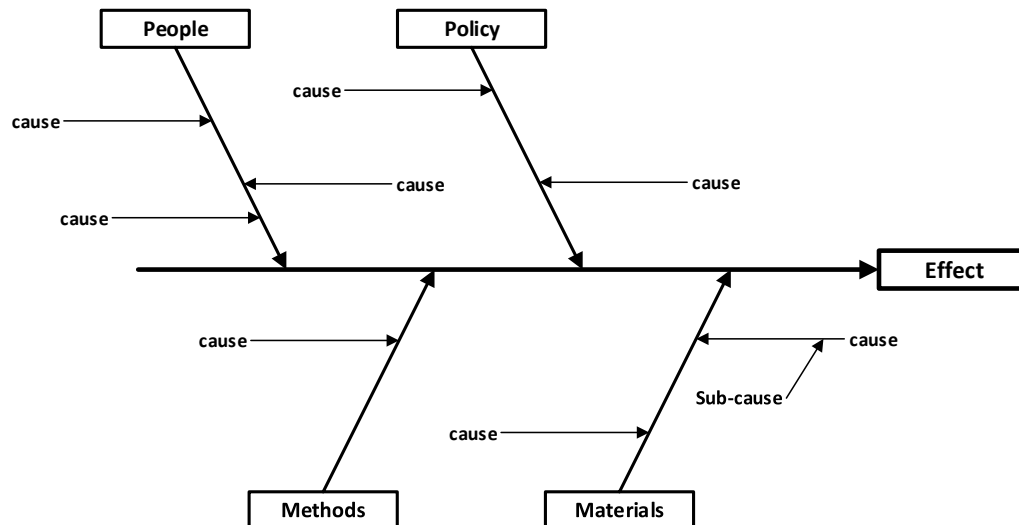


Step 3: Write down key process inputs in the boxes that you drew in Step 2.

³SPC: Statistical Process Control

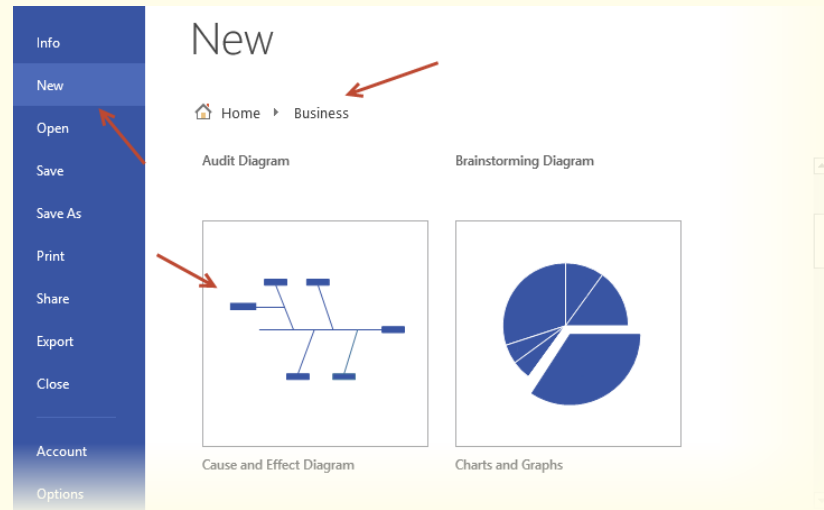


Step 4: From each sub-branch, draw horizontal arrows for the causes. Sub-causes can be added as necessary.

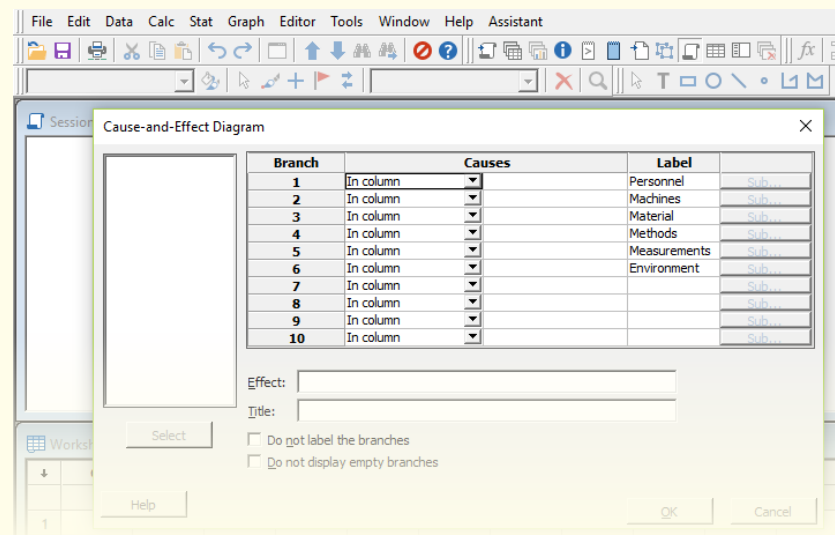


How-To 6.1 (Creating a cause-and-effect diagram in Microsoft Visio 2013)

Open Microsoft Visio and click on *File > New > Business > Cause and Effect Diagram*. You can also access the cause-and-effect diagram option by clicking on *File > New > type **cause and effect** in the search box for templates*. See the snapshot in Figure 6.1 for illustration. Change the default labels as necessary.

Figure 6.1: Options for a cause-and-effect diagram in Microsoft Visio 2013**How-To 6.2 (Creating a cause-and-effect diagram in Minitab 18)**

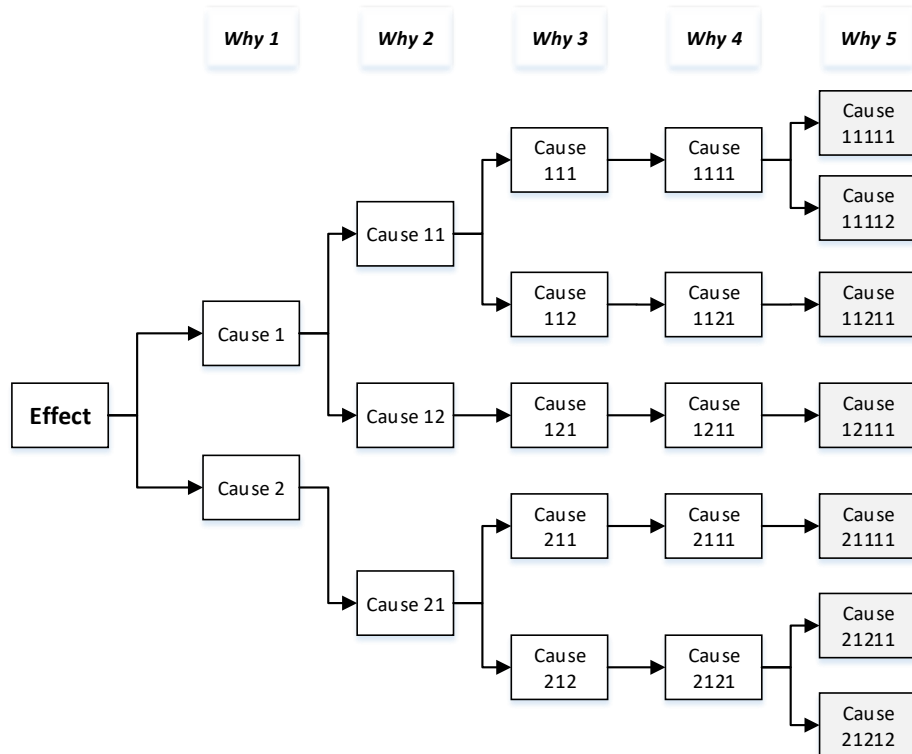
Open Minitab and click on *Stat > Quality Tools > Cause-and-Effect*. See the snapshot in Figure 6.2.

Figure 6.2: Options for a cause-and-effect diagram in Minitab 18**6.4.3 5Whys diagram**

A **5Whys diagram**, as illustrated in Figure 6.3, is a graphical tool that conceptually accomplishes the same goal as the cause-and-effect diagram. One difference is that, unlike in the latter, in the former, we have to keep asking why (e.g., 5 times) until we have identified the atomic cause of the problem. Another difference is that in a 5Whys diagram, we

don't have to organize causes by particular inputs.

Figure 6.3: A general structure of a 5Whys diagram



See an example of the application of a 5Whys diagram in [32].

6.4.4 Why-why diagram

A **Why-Why diagram** is a variation of the 5Whys diagram that we can use to analyze the risk of an adverse event. Unlike the 5Whys diagram, where we ask the causes of an effect, in the Why-Why diagram, we inquire about how an adverse event could happen [55]. The setup of this diagram looks like that of a 5Whys diagram portrayed in Figure 6.3.



How-To 6.3 (Creating a Why diagrams in Microsoft Visio 2013)

To draw a 5Whys diagram, use the basic shapes in Visio and connect boxes using arrows.

6.4.5 Is-Is Not matrix

An **Is-Is Not** matrix, as illustrated in Table 6.1, is a tabular tool that we can employ to evaluate the risks of quality problems by contrasting favorable and unfavorable conditions

of an adverse event [31, 55]. A careful analysis of the differences between the **Is** and the **Is Not** columns can help reveal potential high-risk areas in the process.

Table 6.1: A basic setup of the **Is-Is Not** matrix

Adverse event (specify)	Is	Is Not	Difference
Where did the event happen?	Specify the location where the event occurred	Specify a comparable location where the event did not occur	What is the difference? State a plausible hypothesis.
Who was affected by the event?	Specify the type of the customer affected	Specify the type of the customer not affected	...
How did the event happen?	Specify how the event happened	Specify how the event did not happen (e.g., prevented)	...
When did the event happen?	Specify the time, shift, and date of occurrence	Specify when the event did not happen	...
...

6.5 Prioritizing ideas

6.5.1 Voting techniques

After organizing ideas into major themes or classes, we need to find a way of prioritizing them. Some of the voting techniques that we can apply include:

- **Nominal Group Technique (NGT)**, where the team ranks and prioritizes ideas from the most to the least important [49].
- **Multivoting Technique**, which is a version of NGT where instead of ranking ideas, the team rates them [49]. In practice, multivoting is an iterative process where the team continues to eliminate low rated ideas until only a few useful suggestions remain [55].
- **Q-sort** method, which is another version of NGT where the team ranks and sorts ideas in descending order of importance [55]. A simple algorithm of this technique follows:

Step 1: Have each team member (voter) rank ideas

Step 2: Organize the rankings by *idea* and *voter*.

Step 3: Sum or average the rankings of each idea.

Step 4: Sort ideas in descending order of magnitude.

We demonstrate how to implement the Q-sort method in Example 6.1.

Example 6.1 (Q-sort method)

The manager of health informatics at Starling City Hospital has requested ideas about how to improve the turn-around-time of record requests. Six ideas (A, \dots, F) were generated, and three voters were recruited ($Voter\#1, Voter\#2$, and $Voter\#3$) to help rank these ideas. The Q-sort algorithm was implemented, as follows:

Step 1					Step 2				
Rank	Voter #1	Voter #2	Voter #3		Idea	Voter #1	Voter #2	Voter #3	
6	F	E	D		A	5	3	4	
5	A	B	E		B	2	5	1	
4	C	C	A		C	4	4	3	
3	E	A	C		D	1	2	6	
2	B	D	F		E	3	6	5	
1	D	F	B		F	6	1	2	

Step 3					Step 4				
Idea	Voter #1	Voter #2	Voter #3	Sum	Idea	Voter #1	Voter #2	Voter #3	Sum
A	5	3	4	12	E	3	6	5	14
B	2	5	1	8	A	5	3	4	12
C	4	4	3	11	C	4	4	3	11
D	1	2	6	9	D	1	2	6	9
E	3	6	5	14	F	6	1	2	9
F	6	1	2	9	B	2	5	1	8

In this case, idea **E** has the highest rank and should, therefore, be prioritized for quality improvement efforts of the record request process.

6.5.2 Failure Modes and Effects Analysis (FMEA)

FMEA is a technique that we employ to quantify risk in terms of the factors of the **likelihood**, **detectability**, and **severity** of the failure [43]. We typically rate each factor on a scale of *1 to 5* [55] or *1 to 10* [43]. The implementation of the scale of *1 to 10* looks like this:

1. The **likelihood** of the failure (rated on 1 to 10 scale with 1=not likely and 10 = almost certain)
2. The **detectability** of the failure (rated on 1 to 10 scale with 1 = very likely to detect and 10 = very unlikely to detect).
3. The **severity** of the failure (rated on a 1 to 10 scale with 1 = little impact and 10 = catastrophic impact)

To measure the risk of each potential problem, we calculate its **risk priority number (RPN)**, as follows:

$$RPN = likelihood \times detectability \times severity \quad (6.1)$$

The problem with the highest RPN should be prioritized for improvement or prevention efforts.

Example 6.2 (FMEA)

Metropolis Hospital would like to quantify the risk of potential patient care problems using the FMEA tool. The issues of concern are:

1. The wrong site of surgery
2. Medication error
3. Nosocomial infection
4. Injury

After an in-depth discussion with medical providers and an extensive analysis of historical data, the quality team rated each adverse effect on a scale of 1 to 10, as portrayed in Table 6.2.

Table 6.2: Potential quality risks in patient care

	Severity	Likelihood	Detectability
The Wrong site of surgery	10	1	2
Medication error	8	3	4
Nosocomial infection	6	4	1
Injury	7	5	1

Subsequently, the RPN measure of each event was determined as summarized in Table 6.3.

Table 6.3: RPN measures

	Severity	Likelihood	Detectability	RPN
Medication error	8	3	4	96
Injury	7	5	1	35
Nosocomial infection	6	4	1	24
The Wrong site of surgery	10	1	2	20

Based on the RPN results in Table 6.3, Metropolis Hospital should prioritize the prevention efforts of **medication errors**.

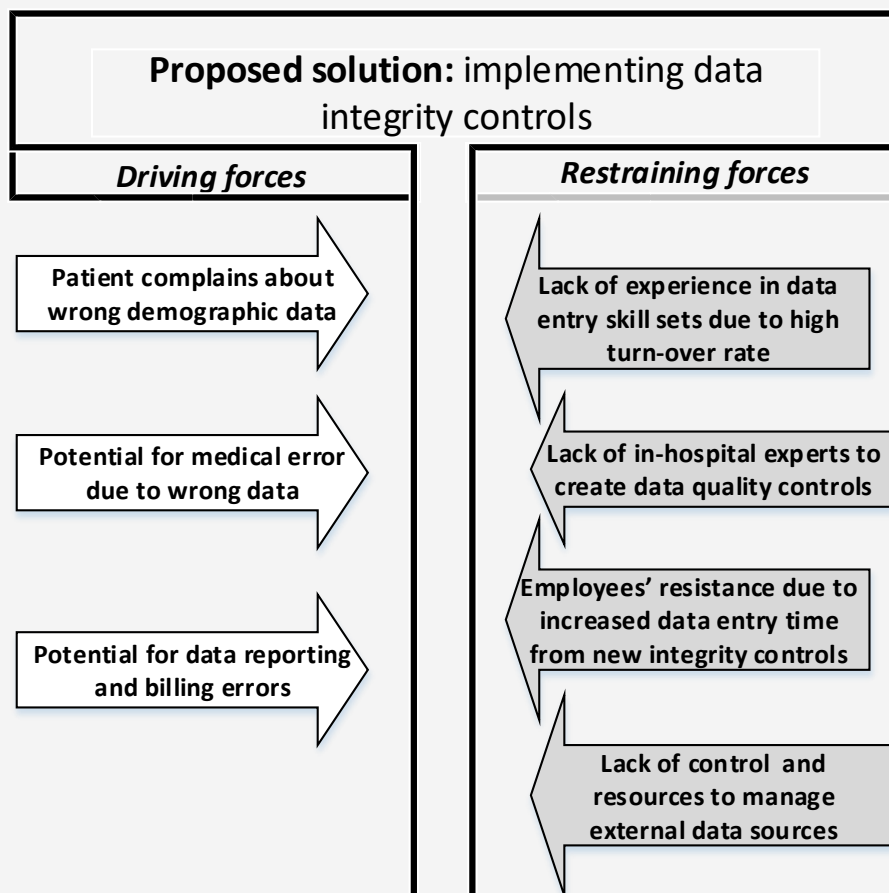
6.5.3 Force-field analysis

Force-field analysis is a technique that is broadly used to evaluate driving and restraining forces of a particular event. In the context of quality improvement, we could apply this tool in the *Analyze* and *Improve* phases of DMAIC to identify potential risks and hindering forces to the proposed solutions. We illustrate how to create a force-field plot in Example 6.3.

Example 6.3 (Force-Field Analysis)

After conducting several data quality audits, the new manager of health informatics at Metropolis Hospital has determined that nearly 80% of all data quality problems are the result of *entry errors*. The manager believes that strong data integrity controls and friendlier user interfaces could fix the problem. So, the manager is suggesting automating many data entry tasks, creating drop-down menus of options, instituting immediate validation of entered data, and enforcing all referential and domain integrity rules. Before implementing these solutions, the manager held brainstorming sessions with employees and experts to request feedback about the proposed changes. Subsequently, the manager organized the received ideas in a force-field analysis plot, as depicted in Figure 6.4.

Figure 6.4: A force-field analysis of data quality problems at Metropolis Hospital



The manager will use the results from this analysis to refine and implement feasible solutions. A pilot implementation, on a small scale, is recommended.

6.6 Check your understanding

1. What visual tool can we use to evaluate a **potential relationship** between two variables?
 - (a) Scatter plot
 - (b) Cause-and-effect diagram
 - (c) Box plot
 - (d) Stem-and-leaf diagram
2. Which tool can we apply **to group similar ideas**?
 - (a) Q-sort
 - (b) Affinity
 - (c) Failure mode
 - (d) Marginal plots
3. When we specify the type of customer who was affected vs. the type of customer who was not affected, we are likely performing
 - (a) an is-is not analysis
 - (b) a cause-and-effect analysis
 - (c) force-field analysis
 - (d) failure mode and effect analysis
4. **Restraining forces** are often discussed in which type of analysis?
 - (a) Cause-and-effect
 - (b) Force-field
 - (c) Failure mode
 - (d) Scatter

5. From a brainstorming session, you have generated the following potential causes of IT events.

recent changes in network/server vendor to save money; some workstations don't have updated anti-virus programs; network switch configuration; bandwidth congestion; shortage of IT staff by 2 programmers; lack of software maintenance by vendor; users lack training/skills to use software properly; lack of data governance policies and procedures; lack of communication among IT staff; some hardware in unsafe locations; legacy computer system; lack of integrity control on data entry; password sharing among users; dependence on technology; no plan B for power failures; outdated business rules.

Organize these ideas on a fishbone diagram. What can you conclude?

6. Following from the previous question, organize the potential causes into, at most, four significant classes using affinity grouping? Justify your classification and discuss the significance of the resulting groups.
7. Consider the problem of medication errors. Consult external sources and organize potential causes using a 5Whys diagram. Explain your work and cite your sources.
8. Following from the previous question, suppose you have concluded that a CPOE system could help mitigate medication errors. Construct a force-field analysis and identify driving and potential hindering forces to your solution.
9. The manager of Health Informatics at Metropolis Hospital has rated potential risks of data quality problems, as illustrated in Table 6.4. What issue should be prioritized, given the risk?

Table 6.4: Potential risks of data quality problems at Metropolis Hospital

	Severity	Likelihood	Detectability
Non-unique entities	6	1	1
Information from the wrong patient	10	3	9
Incomplete non-full fields	9	1	1
Outdated security standards	8	2	6
Inconsistent metadata	3	7	4
Downtime of the EHR system	4	10	1
Typos in the free text fields	7	9	5

10. You have consulted a team of 4 experts to help vote on the best idea for improving the problem of inconsistency of medical information in the EHR system of your hospital. There are 8 possible solutions (A,..., H), and each expert ranked them from best to least, as follows:
 - (a) Expert_1: [H, B, G, F, D, C, A, E]
 - (b) Expert_2: [C, B, H, G, F, D, A, E]
 - (c) Expert_3: [E, A, D, H, C, B, G, F]
 - (d) Expert_4: [E, F, G, H, A, B, C, D]

Rank the solutions using the Q-sort technique. Discuss your results.

MODULE 7

Tools for benchmarking quality

In this module, we review techniques for benchmarking quality by comparing sample or population statistics from different sources in the process. The statistical concepts that we apply include hypothesis testing, the analysis of variance (ANOVA), and the chi-square test. Other tools we discuss include league tables, the analysis of means (ANOM), and funnel charts.

Key concepts and tools: Benchmarking quality; Measuring a process; Gauge analysis; Repeatability & Reproducibility (R&R) study; ANOVA; ANOM; Chi-square test; League tables; Contingency tables; Contingency tables; Report cards; Funnel charts

Major objectives

After studying this module, you will be able to:

- 1. Define key concepts and tools of benchmarking quality*
- 2. Understand the need for measuring and benchmarking quality*
- 3. Discuss the use of quality report cards in health care*
- 4. Identify common statistical tools for comparing sample or population statistics*
- 5. Apply hypothesis testing techniques*
- 6. Implement funnel charts*
- 7. Implement ANOM charts*
- 8. Apply the ANOVA test to analyze the R&R study*
- 9. Compute and interpret Chi-square tests*
- 10. Advocate for benchmarking methods to improve quality in health care*

7.1 Introduction

In this module, we review techniques for “benchmarking quality” by comparing sample or population statistics from different sources in the process. The tools that we discuss include hypothesis tests that, ANOVA, Chi-square tests, ANOM, and funnel charts.

7.2 Statistical background

7.2.1 Confidence interval

In Module 4, we briefly discussed the concepts of sample statistics and point estimators. In this subsection, we extend these ideas into the related topic of **confidence interval**.

Definition 7.1 A **confidence interval** is a range constructed from sample statistics to estimate a probabilistic interval that covers a given population parameter [36]

We create a confidence interval by setting lower and upper confidence limits of a parameter of interest. Here, we only review how to generate the confidence intervals of the population mean μ and variance σ^2 when the population is normally distributed.

Given a lower confidence limit A and an upper confidence limit B , we denote the probability of the population mean μ falling between A and B as follows:

$$P\{A \leq \mu \leq B\} = 1 - \alpha \quad (7.1)$$

where α is the significance level or the error that we are willing to commit. Using α , we calculate the percentage of our **confidence level** this way: $100(1-\alpha)\%$. Equation 7.1 reflects a **two-sided** confidence interval. A **one-sided** confidence interval looks like this:

$$P\{A \leq \mu\} = 1 - \alpha \text{ or } P\{\mu \leq B\} = 1 - \alpha \quad (7.2)$$

The confidence interval of the population mean μ

Assuming normally distributed population, we construct a confidence interval of μ by considering the following two cases:

Case 1: a confidence interval of μ when σ is known

From the notion of CLT in Definition 4.3, we use the standardized value Z to construct the confidence interval of μ this way:

$$\mu = \bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \quad (7.3)$$

where $\alpha/2$ signifies a two-sided confidence interval. In a one-sided confidence interval, we replace $\alpha/2$ with α . We typically construct this confidence interval when the sample size n is large enough to allow for a good point estimation of the population standard deviation (e.g., $n \geq 30$ [36]).

Case 2: a confidence interval of μ when σ is unknown

In this case, we use the t-distribution to determine the confidence interval of μ . Using Equation 4.9, we proceed as follows:

$$\mu = \bar{x} \pm t_{\alpha/2, n-1} \frac{s}{\sqrt{n}} \quad (7.4)$$

where s is the sample standard deviation, and $n - 1$ represents the degrees of freedom. In a one-sided confidence interval, we replace $\alpha/2$ with α . We typically employ this confidence interval technique for smaller sampler sizes (e.g., $n < 30$ [36])

The confidence interval of the population variance σ^2

When the population is normally distributed, and we know neither the mean μ nor the variance σ^2 , we use sample statistics and the Chi-square distribution χ^2 to create a two-sided confidence interval for σ^2 , as follows [43]:

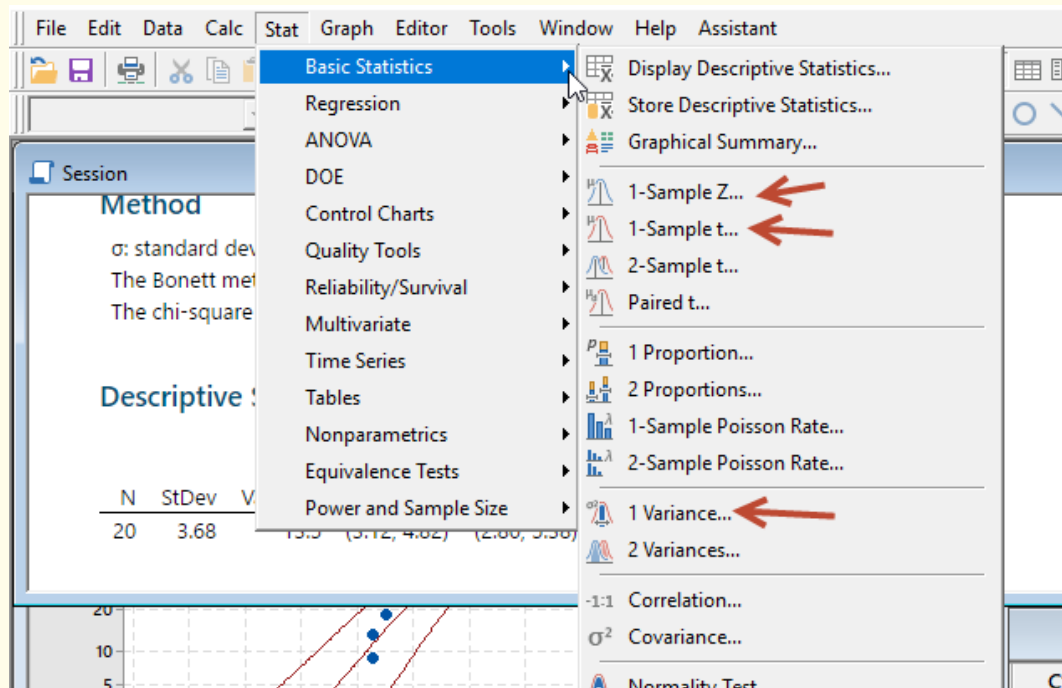
$$\frac{(n-1)s^2}{\chi_{\alpha/2, n-1}^2} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi_{1-\alpha/2, n-1}^2} \quad (7.5)$$

Here, s^2 is the sample variance, and $n - 1$ represents the degrees of freedom. We determine a one-sided confidence interval this way:

$$\sigma^2 \leq \frac{(n-1)s^2}{\chi_{1-\alpha, n-1}^2} \quad (7.6)$$

How-To 7.1 (Confidence intervals of μ and σ^2 in Minitab 18)

Click on *Stats > Basic Statistics*. For the confidence interval of μ when σ is known, choose *1-Sample Z*. For the confidence interval of μ when σ is unknown, choose *1-Sample t*. For the confidence interval of σ^2 , choose *1 Variance*. **Note:** Minitab may give you the confidence interval of the standard deviation, instead of that of the variance. Simply square the given limits to obtain the confidence interval of the variance. See the snapshot in Figure 7.1.

Figure 7.1: Options for the confidence intervals of μ and σ^2 in Minitab 18**How-To 7.2 (Confidence intervals of μ and σ^2 in Excel 2013)**

There is no special option in Excel to calculate confidence intervals, but we can use the built-in functions to estimate the interval of interest. For example, we estimate the confidence interval for the population mean like this:

1. When σ is known, we use the function $= -NORM.S.INV(\alpha/2)$ to determine $Z_{\alpha/2}$. For example, when $\alpha = 0.05$, we find $Z_{0.05/2} = Z_{0.025}$ by $= -NORM.S.INV(0.025) = 1.96$.
2. When σ is unknown, we use the function $= -T.INV(\alpha/2, n - 1)$ to obtain the critical value of the t-distribution. For example, when $\alpha = 0.05$, and $n = 10$, we find the critical value by $= -T.INV(\alpha/2, n - 1) = -T.INV(0.025, 9) = 2.262$.

We can use Excel to construct the confidence interval for the population variance σ^2 by first determining the pertaining values of the Chi-square distribution. For example, we determine $\chi^2_{1-\alpha/2, n-1}$ as $= CHISQ.INV.RT(1 - \alpha/2, n - 1)$. To obtain $\chi^2_{\alpha/2, n-1}$, we use this function: $= CHISQ.INV.RT(\alpha/2, n - 1)$.

Example 7.1 (Confidence Interval in Excel 2013)

Following from Example 4.1, we show how to determine a two-sided 95% confidence interval of μ for the sample on day #1. Given σ is unknown, we determine



the confidence interval this way:

$$\mu = 1.32 \pm 2.77645(0.4/\sqrt{5}) \quad (7.7)$$

$$= (0.82, 1.82) \quad (7.8)$$

The value 2.77645 is the critical value that we obtained by $= T.INV(0.025, 4)$, where $\alpha/2 = 0.025$ and $n - 1 = 5 - 1 = 4$ are degrees of freedom.

7.2.2 Hypothesis testing

We conduct a hypothesis test to evaluate the likelihood of our belief about the value of the population parameter. During hypothesis testing, we make two statements. Our first statement that we call the **null hypothesis** or H_0 expresses our belief that the population parameter equals a particular value or falls within a given range. Our second statement that we call an **alternative hypothesis** or H_1 , opposes the first statement. We **reject** H_0 , when a derived **test statistic** is greater than a given **critical value** or when the **p-value** $< \alpha$. Otherwise, we **fail to reject** H_0 . As before, α is the significance level, or simply the error that we are willing to commit.

Hypothesis test for the process mean

We start by assuming that $\mu = \mu_0$, where μ_0 is the conjectured value. We formulate a **two-tailed** or two-sided hypothesis test as follows:

$$H_0 : \mu = \mu_0 \quad (7.9)$$

$$H_1 : \mu \neq \mu_0 \quad (7.10)$$

Alternatively, we can create a **one-tailed** or one-sided test this way:

$$H_0 : \mu = \mu_0 \quad (7.11)$$

$$H_1 : \mu < \mu_0 \quad (7.12)$$

or as:

$$H_0 : \mu = \mu_0 \quad (7.13)$$

$$H_1 : \mu > \mu_0 \quad (7.14)$$

Next, we find and compare the **test statistic** Z_0 to the corresponding **critical value** Z_α . In a normally distributed process where σ is **known**, we refer to our hypothesis test as the **Z-test**. We obtain the test statistic Z_0 as follows:

$$Z_0 = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \quad (7.15)$$

The critical value is given by $Z_{\alpha/2}$ in a two-sided test or Z_α in a one-sided test. We reject H_0 when $|Z_0| > Z_{\alpha/2}$ in a two-sided test or when $|Z_0| > Z_\alpha$ in a one-sided test. We use $|\cdot|$

symbols to denote the absolute value function. When the population standard deviation σ is **unknown**, we conduct a **t-test**, instead of a Z-test, and obtain the test statistic t_0 this way:

$$t_0 = \frac{\bar{x} - \mu}{s/\sqrt{n}} \quad (7.16)$$

Given the degrees of freedom $n-1$, the critical value is obtained by $t_{\alpha/2, n-1}$ in a two-sided test or $t_{\alpha, n-1}$ in a one-sided test. In a two-sided test, we reject H_0 when $|t_0| > t_{\alpha/2, n-1}$. In a one-sided test, we reject H_0 when $|t_0| > t_{\alpha, n-1}$.

Hypothesis test for the process variance

We structure a hypothesis test for the process variance like this:

$$H_0 : \sigma^2 = \sigma_0^2 \quad (7.17)$$

$$H_1 : \sigma^2 \neq \sigma_0^2 \quad (7.18)$$

where σ_0^2 is the variance that we have conjectured. Alternatively, we could set $H_1 : \sigma^2 > \sigma_0^2$ or $H_1 : \sigma^2 < \sigma_0^2$. Assuming normal distribution and unknown population mean and variance, we use the Chi-square distribution to obtain the test statistic, χ_0^2 , as follows:

$$\chi_0^2 = \frac{(n-1)s^2}{\sigma_0^2} \quad (7.19)$$

We recall that s^2 is the sample variance, and $n-1$ represents the degrees of freedom. The two-sided critical value is given by $\chi_{\alpha/2, n-1}^2$. We reject the null hypothesis when $\chi_0^2 > \chi_{\alpha/2, n-1}^2$ or $\chi_0^2 < \chi_{1-\alpha/2, n-1}^2$. The critical value in a one-sided test is $\chi_{\alpha, n-1}^2$. For $H_1 : \sigma^2 > \sigma_0^2$, we reject the null hypothesis when $\chi_0^2 > \chi_{\alpha, n-1}^2$. For $H_1 : \sigma^2 < \sigma_0^2$, we reject the null hypothesis when $\chi_0^2 < \chi_{1-\alpha, n-1}^2$ [43].

7.2.3 P-value

We have briefly introduced p-value before as a probability measure of a hypothesis test. We formally define this quantity next.

Definition 7.2 (P-value)

p-value is the probability that one would observe a test statistic that was greater or equal to the given critical value, if the null hypothesis were true [36]

We recall that the critical value is a statistic that we construct using the confidence level α . When **p-value** $< \alpha$, we reject the null hypothesis. Most statistical packages, such as Minitab, provide us with p-values automatically when we run hypothesis tests. In a normal distribution, we calculate p-values as follows [43]:

$$1. H_1 : \mu \neq \mu_0, \text{ p-value} = 2(1 - \Phi(|Z_0|))$$

$$2. H_1 : \mu > \mu_0, \text{p-value} = 1 - \Phi(Z_0)$$

$$3. H_1 : \mu < \mu_0, \text{p-value} = \Phi(Z_0)$$

Z_0 is the test statistic that we obtain per Equation 7.15.

7.2.4 The type of error

While testing our hypothesis, we are likely to make two types of errors:

Type I error (α): we reject H_0 when H_0 is true.

Type II error (β): we fail to reject H_0 when H_0 is false.

Operationally speaking, the α error is equivalent to erroneously rejecting the quality of a good product or service. In control chart applications, this type of error is also referred to as **false alarm** and signifies the chances of observing erroneous out-of-control signals. The β error is equivalent to mistakenly delivering a bad product or service to the customer [43]. We choose α to indicate the error probability that we are willing to tolerate. When our process is normally distributed, we typically set $\alpha = 0.05$ and determine β this way [43]:

$$\beta = \Phi\left(Z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right) - \Phi\left(-Z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right) \quad (7.20)$$

Here, δ is the difference between the true process parameter and the hypothesized parameter. For example in the context of the mean, $\delta = \mu_1 - \mu_0$, where μ_1 is the true mean and μ_0 is the hypothesized mean. From the β error, we obtain the **power of our statistical test** as like this:

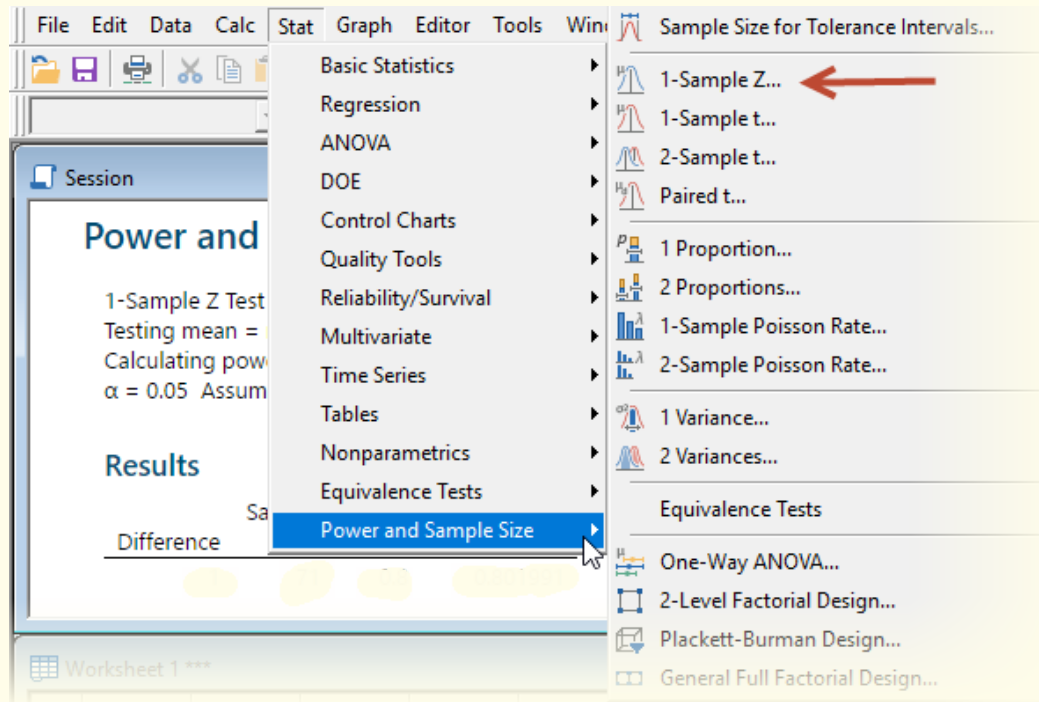
$$\text{Power} = 1 - \beta \quad (7.21)$$

This power test expresses the probability of rejecting the null hypothesis correctly. To increase the power test, we increase the sample size n , which in turn decreases β per Equation 7.20. The **operating-characteristic (OC)** curves can help us choose the proper sample size to satisfy a particular value of β [43]. Minitab has an option for obtaining the appropriate sample size n when the desired power test is provided. Minitab instructions for the power of 1-sample Z test are shown in 7.3.

How-To 7.3 (Power for 1-sample Z test in Minitab 18)

Click on *Stat > Power and Sample Size > select the test of interest*. For example, to determine the sample size in 1-Sample Z Test, leave the sample size field blank, and specify the difference that you want to detect δ . You also need to specify the population standard deviation and the desired power value. See the snapshot in Figure 7.2.



Figure 7.2: Power and Sample Size for 1-Sample Z Test in Minitab 18

7.3 Comparing the means of two independent samples

We can use hypothesis testing to check whether the means of two **independent** samples were drawn from the same population or two similarly distributed populations. To test whether the population means, μ_1 and μ_2 , are equal, we obtain two samples of sizes n_1 and n_2 and means \bar{x}_1 and \bar{x}_2 , respectively. We formulate the null hypothesis like this:

$$H_0 : \mu_1 = \mu_2 \quad (7.22)$$

We frame alternative hypotheses as $H_1 : \mu_1 \neq \mu_2$, $H_1 : \mu_1 > \mu_2$, or $H_1 : \mu_1 < \mu_2$. Assuming a normally distributed population with the known variance σ^2 , we apply the Z-test and obtain the **test statistic** as follows:

$$Z_0 = \frac{\bar{x}_1 - \bar{x}_2}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (7.23)$$

If the variance is unknown, we apply the t-test and obtain the **test statistic** like this:

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (7.24)$$

Here, s_p is the square root of the pooled variance s_p^2 that is computed this way:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \quad (7.25)$$

where $n_1 + n_2 - 2$ represents the degrees of freedom from two samples [36]. We obtain critical values and conclude about the null hypothesis as before.

7.4 Comparing the means of more than two independent samples

Under the assumptions of the normal distribution, independence, and equal variances, we can test the following hypothesis [36]:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_m \quad (7.26)$$

$$H_1 : \text{the means are not all equal} \quad (7.27)$$

Here, m is the index of the samples. Using the Analysis of Variance (ANOVA) test, we obtain the following **test statistic**:

$$F_0 = \frac{MS_{treatments}}{MS_{error}} \quad (7.28)$$

where *treatments* symbolizes independent samples, $MS_{treatments}$ is the mean square of treatments, and MS_{error} is the mean square of the error. Our **critical value** originates from the F-distribution and looks like this:

$$F_{\alpha, m-1, m(n-1)} \quad (7.29)$$

If $F_0 > F_{\alpha, m-1, m(n-1)}$, we reject the null hypothesis. We will return to the topic of ANOVA in Section 7.8.

7.5 Comparing the means of two dependent samples

When two samples are **dependent**, we conduct hypothesis testing using the **paired t-test**. Our **test statistic** looks like this:

$$t_0 = \frac{\bar{d}}{s_d / \sqrt{n}} \quad (7.30)$$

where \bar{d} , the mean of the difference between two samples (each with size n), is given by:

$$\bar{d} = \frac{1}{n} \sum_{j=1}^n d_j \quad (7.31)$$

From Equation 7.30, s_d is the standard deviation of the differences, which is obtained by taking the square root of the variance s_d^2 , given by:

$$s_d^2 = \frac{1}{n-1} \sum_{j=1}^n (d_j - \bar{d})^2 \quad (7.32)$$

As before, the critical value is given by $t_{\alpha/2, n-1}$ in a two-sided test or $t_{\alpha, n-1}$ in a one-sided test. In a two-sided test, we reject H_0 when $|t_0| > t_{\alpha/2, n-1}$. In the one-sided test, we reject H_0 when $|t_0| > t_{\alpha, n-1}$.

7.6 Comparing the means of more than two dependent samples

Standard statistical software such as Minitab have options to compare the means of more than two dependent samples, such as when one repeats measures on single or double factors. See Minitab instructions in How-To 7.6 about how to analyze a design of a single factor with repeated measures.

How-To 7.4 (Z-test and t-test in Excel 2013)

You can run a t-test in Excel by using this function: $=T.TEST(sample_1, sample_2, tails, type)$. In the *tails* option, you specify whether the test is one or two-tailed. In the *type* option, you specify whether you are running a paired t-test (1), a t-test with equal variances (2), or a t-test with unequal variances (3). You can also run a t-test via the **Data Analysis add-in**. For that, open the *Data Analysis* add-in, as shown in the snapshot portrayed in Figure 7.3 and select the appropriate test (see the snapshot in Figure 7.4).



Figure 7.3: Excel 2013 Data Analysis add-in option

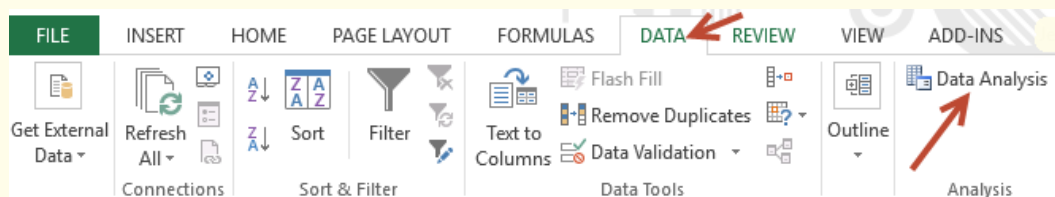
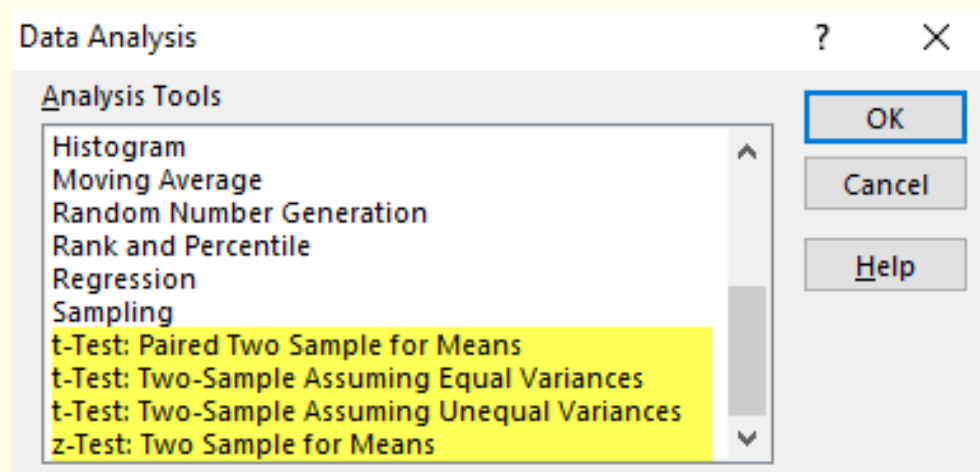


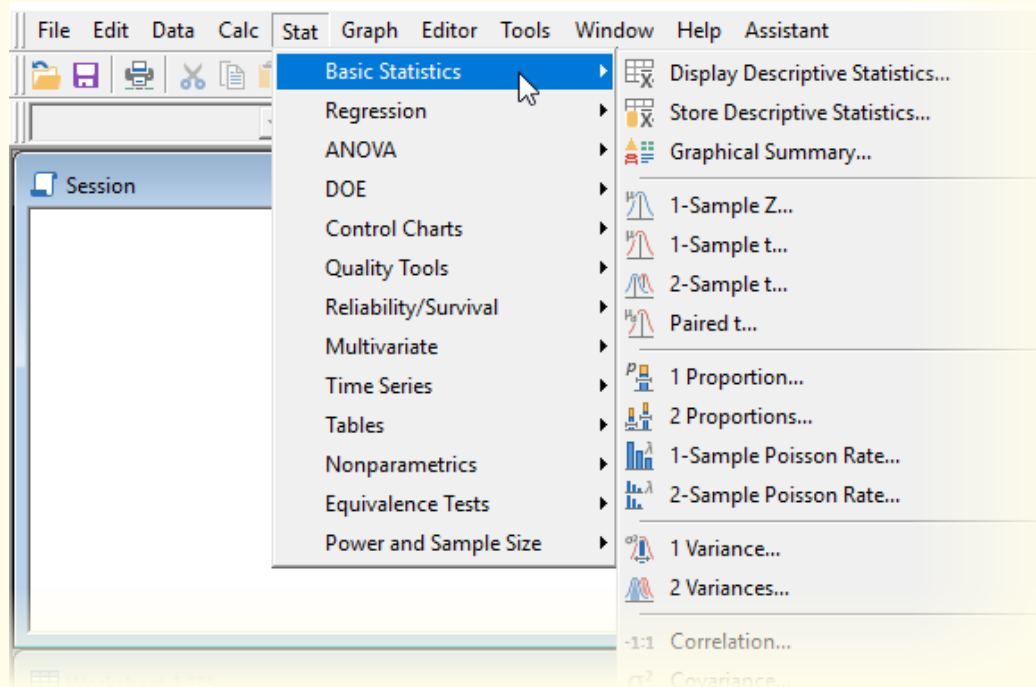
Figure 7.4: Z-test and t-test via Data Analysis add-in option in Excel 2013



If the Data Analysis add-in is not loaded, you can add it by clicking on *File > Options > Add-ins > Analysis ToolPak > Manage: Excel Add-in > Check the box of Analysis ToolPak > OK*.

How-To 7.5 (Z-test and t-test in Minitab 18)

To conduct a Z-test or a t-test in Minitab 18, click on *Stat>Basic Statistics* and select the appropriate test. See the snapshot portrayed in Figure 7.5.

Figure 7.5: Z-test and t-test options in Minitab 18**How-To 7.6 (Single factor with repeated measures in Minitab 18)**

To conduct an analysis of a single factor design with repeated measures in Minitab 18, click on *Stat > ANOVA > General Linear Model > Fit General Linear Model > Select Factors > Select Responses > Click on Random/Nest > Factor Type > Random > OK*.

Example 7.2 (Hypothesis testing)

This example follows the case study and data presented in Example 4.1.

1. For demonstrations purposes, we conduct a hypothesis test to see if sample # 1 and sample #2 in Table 4.1 were drawn from the same population. We state our hypothesis as follows:

$$H_0 : \mu_1 = \mu_2 \quad (7.33)$$

$$H_1 : \mu_1 \neq \mu_2 \quad (7.34)$$

At $\alpha = 0.05$ and unknown variance, we use Equation 7.24 to obtain the test

statistic t_0 this way:

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (7.35)$$

$$= \frac{1.32 - 2.61}{0.36 \sqrt{\frac{1}{5} + \frac{1}{5}}} \quad (7.36)$$

$$= -5.61666444 \quad (7.37)$$

We use Equation 7.25 to compute the value of s_p as follows:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \quad (7.38)$$

$$= \frac{(5 - 1)0.161^2 + (5 - 1)0.103^2}{5 + 5 - 2} \quad (7.39)$$

$$= 0.131875 \quad (7.40)$$

So, $s_p = \sqrt{s_p^2} = \sqrt{0.131875} = 0.363145976 \approx 0.36$. With $n_1 + n_2 - 2 = 5 + 5 - 2 = 8$ and $\alpha/2 = 0.05/2 = 0.025$. We obtain the critical value $t_{0.025,8}$ in Excel using $=T.INV(0.025,8) = 2.306004135$. Since $|t_0| > 2.306004135$, we reject the null hypothesis. We could have reached the same conclusion by computing the p-value in Excel as follows:

$$p\text{-value} = T.TEST(B3:B4, B4:B4, 2, 2) \quad (7.41)$$

$$= 0.0005 \quad (7.42)$$

Since $p\text{-value} < \alpha = 0.05$, we once again reject the null hypothesis.

7.7 Comparing variances of two independent samples

We hypothesize that two samples (sample 1 and sample 2), with variances s_1^2 and s_2^2 , were drawn from two independent populations with equal variances, σ_1^2 and σ_2^2 . We create a hypothesis test to validate our conjecture as follows:

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad (7.43)$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2 \quad (7.44)$$

Using F-distribution, we obtain the **test statistic** this way:

$$F_0 = \frac{s_1^2}{s_2^2} \quad (7.45)$$

The lower **critical value** is given by $F_{1-\alpha/2, n_1-1, n_2-1}$ whereas the upper critical value is given by $F_{\alpha/2, n_1-1, n_2-1}$. Here, $n_1 - 1$ and $n_2 - 1$ are degrees of freedom of sample 1 and

sample 2, respectively. If $F_0 > F_{\alpha/2, n_1-1, n_2-1}$ or $F_0 < F_{1-\alpha/2, n_1-1, n_2-1}$ we reject the null hypothesis. To perform a one-sided test, we formulate H_1 and F_0 statistics this way:

$$H_1 : \sigma_1^2 > \sigma_2^2, \quad F_0 = s_1^2/s_2^2 \quad (7.46)$$

$$H_1 : \sigma_1^2 < \sigma_2^2, \quad F_0 = s_2^2/s_1^2 \quad (7.47)$$

For the case of $H_1 : \sigma_1^2 > \sigma_2^2$, we reject the null hypothesis when $F_0 > F_{\alpha, n_2-1, n_1-1}$. For the case of $H_1 : \sigma_1^2 < \sigma_2^2$, we reject the null hypothesis when $F_0 < F_{\alpha, n_1-1, n_2-1}$.

How-To 7.7 (F-test in Excel 2013)



To calculate the critical value of F_{α, n_1-1, n_2-1} in Excel 2013, use this function = *F.INV.RT*($\alpha, n_1 - 1, n_2 - 1$). You can estimate the p-value by = *F.DIST.RT*($F_0, n_1 - 1, n_2 - 1$). You can also conduct the F-test using the **Data Analysis add-in** under the Data tab. Choose the option of *F-test Two-Sample for Variances*.

How-To 7.8 (F-test in Minitab 18)



Click on *Stat > Basic Statistics > 2 Variances*. Review the snapshot in Figure 7.1. **Note** that Minitab runs the test on equal standard deviations, but this will not alter your conclusion.

7.8 ANOVA

ANOVA is a powerful statistical test that we use to examine the equality of the means in more than two independent samples. We assume that the samples are normally distributed with equal variances. The two types of the ANOVA test that we consider here are **one-way ANOVA** and **two-way ANOVA**. In both cases, we work with the sum of squares (*SS*) to formulate the statistical tests of interest.

7.8.1 One-way ANOVA

In a one-way ANOVA test, we obtain the total sum of squares (SS_{total}) in the data as follows:

$$SS_{total} = SS_{treatments} + SS_{error} \quad (7.48)$$

The word *treatments* is used to denote independent samples. The $SS_{treatments}$ term represents the **between** samples variation, whereas the SS_{error} term symbolizes the **within** samples variation. Let's assume that we have a number of samples, each with

size n . We determine the corresponding sum of squares as follows:

$$SS_{total} = \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2 \quad (7.49)$$

$$SS_{treatments} = n \sum_{i=1}^a (\bar{y}_{i.} - \bar{y}_{..})^2 \quad (7.50)$$

$$SS_{error} = \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{j.})^2 \quad (7.51)$$

where y_{ij} is the entry in row i and column j , $\bar{y}_{..}$ is the overall mean of the data, $\bar{y}_{i.}$ is the mean of row i , and $\bar{y}_{j.}$ is the mean of column j . The degrees of freedom (df) are given by:

$$df_{total} = an - 1 \quad (7.52)$$

$$df_{treatments} = a - 1 \quad (7.53)$$

$$df_{error} = a(n - 1) \quad (7.54)$$

It follows that:

$$df_{total} = df_{treatments} + df_{error} \quad (7.55)$$

Using the sum of squares and the degrees of freedom, we obtain the mean squares (MS) this way:

$$MS_{treatments} = \frac{SS_{treatments}}{df_{treatments}} \quad (7.56)$$

$$MS_{error} = \frac{SS_{error}}{df_{error}} \quad (7.57)$$

The $MS_{treatments}$ term estimates the variance in treatments. The MS_{error} term is the unbiased estimator of the variance of the model residuals in Equation 7.48 [43]. We use the ratio between $MS_{treatments}$ and MS_{error} to obtain the following **F-test statistic**:

$$F_0 = \frac{MS_{treatments}}{MS_{error}} \quad (7.58)$$

A one-tailed **critical value** is given by:

$$F_{\alpha, df_{treatments}, df_{error}} \equiv F_{\alpha, a-1, a(n-1)} \quad (7.59)$$

At the significance level α , we reject the null hypothesis when $F_0 > F_{\alpha, a-1, a(n-1)}$. Our conclusion implies that at least one mean is different. The ANOVA test does not identify the different mean in question.

How-To 7.9 (P-value for the F-test in Excel 2013)

To estimate the p-value for the F-test in Equation 7.58, use = $F.DIST.RT(F_0, df_{treatment}, df_{error})$.

**7.8.2 Two-way ANOVA**

In a two-way ANOVA test, we add a blocking variable $SS_{blocking}$ and SS_{total} becomes:

$$SS_{total} = SS_{treatments} + SS_{blocking} + SS_{error} \quad (7.60)$$

The $SS_{blocking}$ term represents different blocks or conditions under which the sample data were collected. If data were arranged in a table format, we could view **treatments** as independent rows and **blocks** as independent columns. Given a number of treatments and b number of blocks, we formulate the sum of squares of a two-way ANOVA this way:

$$SS_{total} = \sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \bar{y}_{..})^2 \quad (7.61)$$

$$SS_{treatments} = b \sum_{i=1}^a (\bar{y}_{i.} - \bar{y}_{..})^2 \quad (7.62)$$

$$SS_{blocking} = a \sum_{j=1}^b (\bar{y}_{.j} - \bar{y}_{..})^2 \quad (7.63)$$

$$SS_{error} = \sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2 \quad (7.64)$$

$$\equiv SS_{total} - SS_{treatments} - SS_{blocking} \quad (7.65)$$

As earlier, we can view y_{ij} as the entry in row i and column j , $\bar{y}_{..}$ is the overall mean of the data, $\bar{y}_{i.}$ is the mean of row i , and $\bar{y}_{.j}$ is the mean of column or block j . To obtain the degrees of freedom (df), we proceed as follows:

$$df_{total} = ab - 1 \quad (7.66)$$

$$df_{treatments} = (a - 1) \quad (7.67)$$

$$df_{blocking} = (b - 1) \quad (7.68)$$

$$df_{error} = (a - 1)(b - 1) \quad (7.69)$$

where

$$df_{total} = df_{treatments} + df_{blocking} + df_{error} \quad (7.70)$$

We summarize the two-way mean squares and the corresponding F_0 statistics in Table 7.1.

Table 7.1: Mean Squares and the F_0 statistics of a two-way ANOVA test

Source of variation	Mean Square	F_0
Treatments	$MS_{treatments} = \frac{SS_{treatments}}{df_{treatments}}$	$\frac{MS_{treatments}}{MS_{error}}$
Blocks	$MS_{blocking} = \frac{SS_{blocking}}{df_{blocking}}$	$\frac{MS_{blocking}}{MS_{error}}$
Error	$MS_{error} = \frac{SS_{error}}{df_{error}}$	

Just like in a one-way ANOVA test, we draw conclusions about our hypothesis by comparing the F_0 statistic to the *critical value*. We could also use p-values to interpret the significance of the test.

In some applications such as in designed experiments and gauge analysis, we can further decompose SS_{total} to include the term representing the **interaction** between *treatments* and the *blocking* variables. With this new component, SS_{total} becomes:

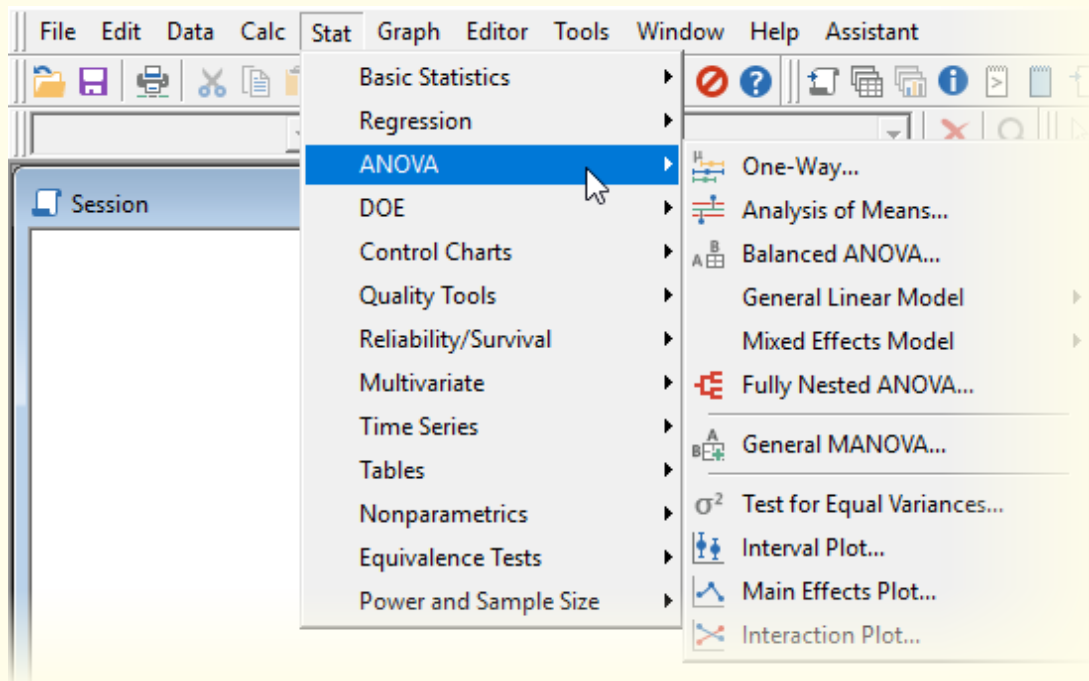
$$SS_{total} = SS_{treatments} + SS_{blocking} + SS_{interaction} + SS_{error} \quad (7.71)$$

From the model in Equation 7.71, the degrees of freedom of the interaction term are given by $(a-1)(b-1)$. We obtain the test statistic by taking the ratio between $MS_{interaction}$ and MS_{error} . The conclusion about the null hypothesis would be as before.

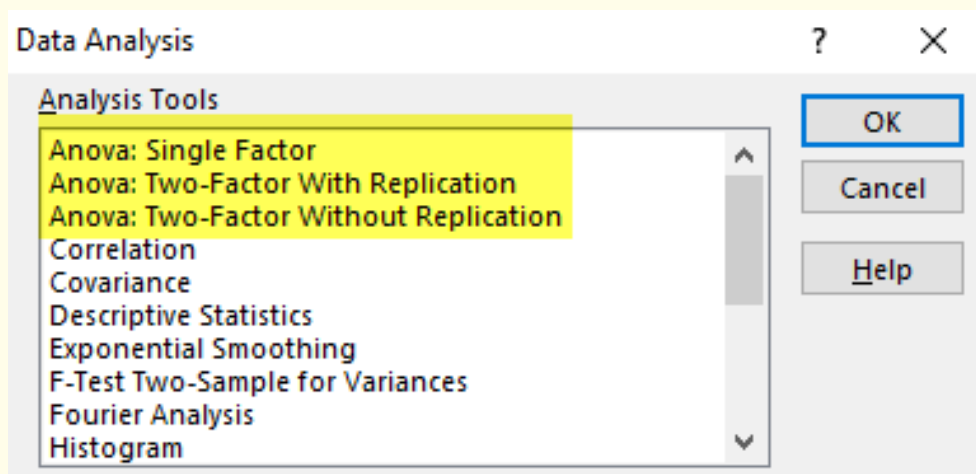
How-To 7.10 (ANOVA test in Minitab 18)

To run the ANOVA test in Minitab 18, click on *Stat > ANOVA*. For a one-way ANOVA test, click on *One-Way*. For a two-way ANOVA test, click on the *General Linear Model*. See the snapshot in Figure 7.6.



Figure 7.6: Options for the ANOVA test in Minitab 18**How-To 7.11 (ANOVA test in Excel 2013)**

Open the *Data Analysis* add-in. For a one-way ANOVA test, use the option of *Anova: Single Factor*. For a two-way ANOVA test without replication, use *Anova: Two-Factor Without Replication*. When the sample data has replications, use *Anova: Two-Factor With Replication*. See the snapshot in Figure 7.7.

**Figure 7.7:** Options for the ANOVA test in Excel 2013 using the Data Analysis add-in

Example 7.3 (ANOVA test and a linear regression model)

The operating room (OR) manager at Metropolis Hospital would like to test whether there is a statistical difference in the OR time for the *Laparoscopic Cholecystectomy* procedures by the physician. Additionally, the manager wants to assess whether the OR location is a significant factor in the regression model since some of the ORs still use old information technology (IT). The manager has hypothesized that there is no difference in the average duration of the procedure either by the physicians or the OR location.

Table 7.2: Averages of OR time for Laparoscopic Cholecystectomy procedures in minutes, Metropolis Hospital, 2018

	PhysicianA	PhysicianB	PhysicianC	PhysicianD	PhysicianE
3N	68	61	58	58	38
3S	72	69	59	63	41
4E	59	64	73	80	48
4W	75	79	77	70	60

As displayed in Table 7.2, the physicians under study are **PhysicianA**, **PhysicianB**, **PhysicianC**, **PhysicianD**, and **PhysicianE**. For OR locations, the manager is evaluating **3N**, **3S**, **4E**, and **4W**.

One-way ANOVA test results

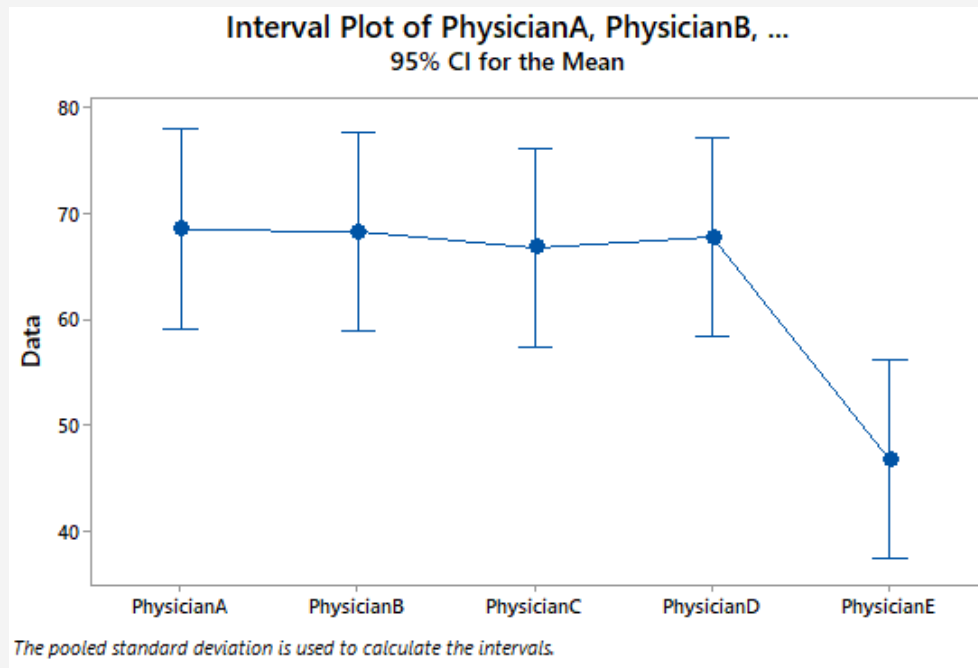
Table 7.3 portrays one-way ANOVA test results for physicians. We obtained these results using the Data Analysis add-in in Excel.

Table 7.3: One-Way ANOVA test for physicians

Anova: Single Factor						
SUMMARY						
Groups	Count	Sum	Average	Variance		
PhysicianA	4	274	68.5	48.33333		
PhysicianB	4	273	68.25	62.25		
PhysicianC	4	267	66.75	93.58333		
PhysicianD	4	271	67.75	90.91667		
PhysicianE	4	187	46.75	95.58333		
ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	1426.8	4	356.7	4.565273	0.013039	3.055568
Within Groups	1172	15	78.13333			
Total	2598.8	19				

From the results in Table 7.3, we notice that the p-value of 0.01 is less than the significance level $\alpha = 0.05$. Besides, we can see that $F > F_{crit}$. So, we reject the null hypothesis that the mean of the operative time among physicians is the same. Minitab yields similar statistical results with an added interval plot displayed in Figure 7.8.

Figure 7.8: A one-way ANOVA interval plot for physicians in Minitab 18

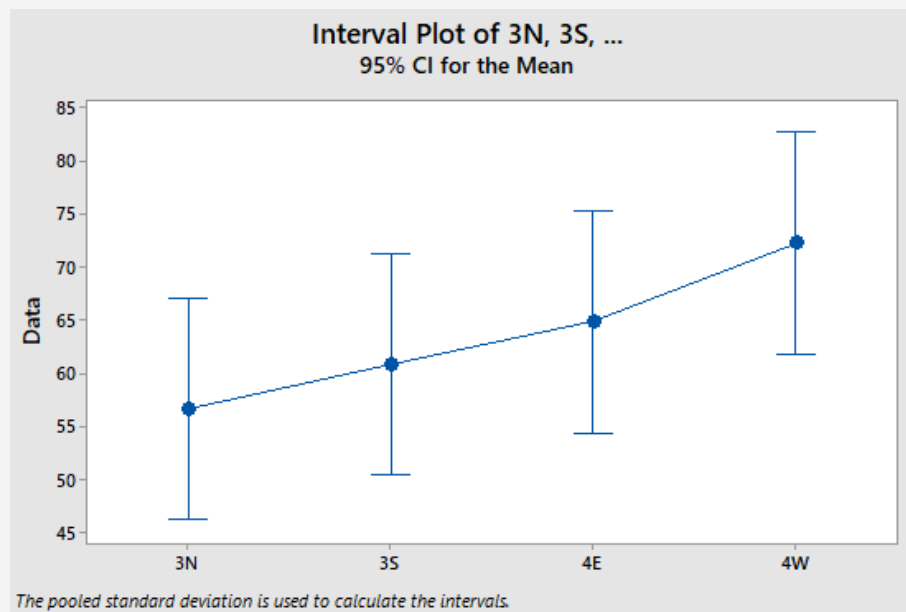


From Figure 7.8, we can deduce that the operative time of PhysicianE is different from that of other physicians, which probably is what has made the statistical test significant. We can also run a one-way ANOVA test by location. Table 7.4 presents results from the Data Analysis add-in in Excel.

Table 7.4: One-Way ANOVA test results for locations

Anova: Single Factor						
SUMMARY						
Groups	Count	Sum	Average	Variance		
3N	5	283	56.6	124.8		
3S	5	304	60.8	148.2		
4E	5	324	64.8	153.7		
4W	5	361	72.2	57.7		
ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	661.2	3	220.4	1.819983	0.184121	3.238872
Within Groups	1937.6	16	121.1			
Total	2598.8	19				

Given that the p-value in Table 7.4 is greater than the significance level α , we fail to reject the null hypothesis. The corresponding Minitab interval plot is portrayed in Figure 7.9

Figure 7.9: A one-way ANOVA interval plot for locations in Minitab 18

Next, we run a two-way ANOVA test to combine both the variables of physicians and locations into one model. This approach will allow us to assess whether there

is a difference in OR time by the physician, given location and vice-versa.

Two-way ANOVA results

Table 7.5 portrays two-way ANOVA test results from the Data Analysis add-in in Excel.

Table 7.5: Excel Two-Way ANOVA test for physicians and locations

Anova: Two-Factor Without Replication				
SUMMARY	Count	Sum	Average	Variance
3N	5	283	56.6	124.8
3S	5	304	60.8	148.2
4E	5	324	64.8	153.7
4W	5	361	72.2	57.7
PhysicianA	4	274	68.5	48.33333
PhysicianB	4	273	68.25	62.25
PhysicianC	4	267	66.75	93.58333
PhysicianD	4	271	67.75	90.91667
PhysicianE	4	187	46.75	95.58333

ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Rows	661.2	3	220.4	5.17776	0.015894	3.490295
Columns	1426.8	4	356.7	8.379796	0.001819	3.259167
Error	510.8	12	42.56667			
Total	2598.8	19				

Based on the p-values in Table 7.5, we conclude that both the physician and location are statistically significant. This time, we reject the hypotheses that the OR time is the same by both physician and location. We recall that the one-way ANOVA test previously indicated no significant difference in the OR time by the location. What changed in the two-way ANOVA test is that the sum of square error decreased after combining both variables, which in turn contributed to the increase in the test statistics of both variables. As we have discussed before, every time we add variables to the model, the sum of the square of the error will decrease. So, we must calculate $R^2_{adjusted}$, instead of simply R^2 , to ascertain the goodness-of-fit of our model. We

use Equation 5.21 and information from Table 7.5 to determine $R^2_{adjusted}$, as follows:

$$R^2_{adjusted} = 1 - \frac{SS_{error}/(n-p)}{SS_{total}/(n-1)} \quad (7.72)$$

$$= 1 - \frac{510.8/(12)}{2598.8/(19)} \quad (7.73)$$

$$= 0.688792263 \quad (7.74)$$

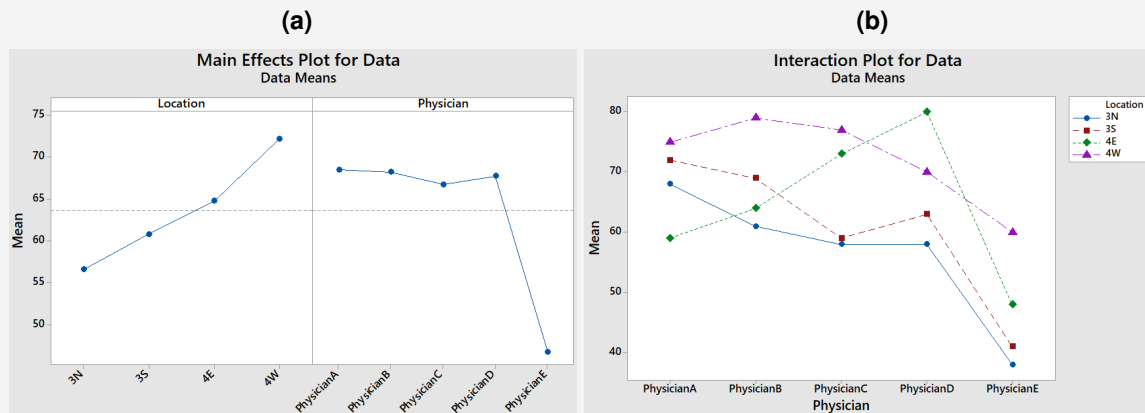
In other words, we can say about 69% of the variation in the OR time can be explained by the variation in both the physician and location variables.

To run a two-way ANOVA test in Minitab, we rearrange the data from Table 7.3 as depicted in Table 7.6. Then, we fit a General Linear Model, as indicated in **How-To 7.10**.

Table 7.6: Reorganized data for a two-way ANOVA test in Minitab 18

Location	Physician	Data
3N	PhysicianA	68
3N	PhysicianB	61
3N	PhysicianC	58
3N	PhysicianD	58
3N	PhysicianE	38
3S	PhysicianA	72
3S	PhysicianB	69
3S	PhysicianC	59
3S	PhysicianD	63
3S	PhysicianE	41
4E	PhysicianA	59
4E	PhysicianB	64
4E	PhysicianC	73
4E	PhysicianD	80
4E	PhysicianE	48
4W	PhysicianA	75
4W	PhysicianB	79
4W	PhysicianC	77
4W	PhysicianD	70
4W	PhysicianE	60

Minitab yields similar ANOVA test results. Furthermore, Minitab provides us with several graphical plots, as portrayed in Figure 7.10. The main effect plot in Subfigure 7.10a suggests that OR time is at the lowest level when 3N is used. This subfigure also indicates that PhysicianE utilizes the least amount of OR time. The interaction plot in Subfigure 7.10b shows variability in the means of operative times by physician and location. From this subfigure, we deduce that 3N yields the lowest operative time for most physicians except physicianA. Additionally, the lowest operative time occurs when PhysicianE uses 3N. The greatest variability in operative time seems to happen when physicians use 4E.

Figure 7.10: Main effects and interaction plots

Linear regression model

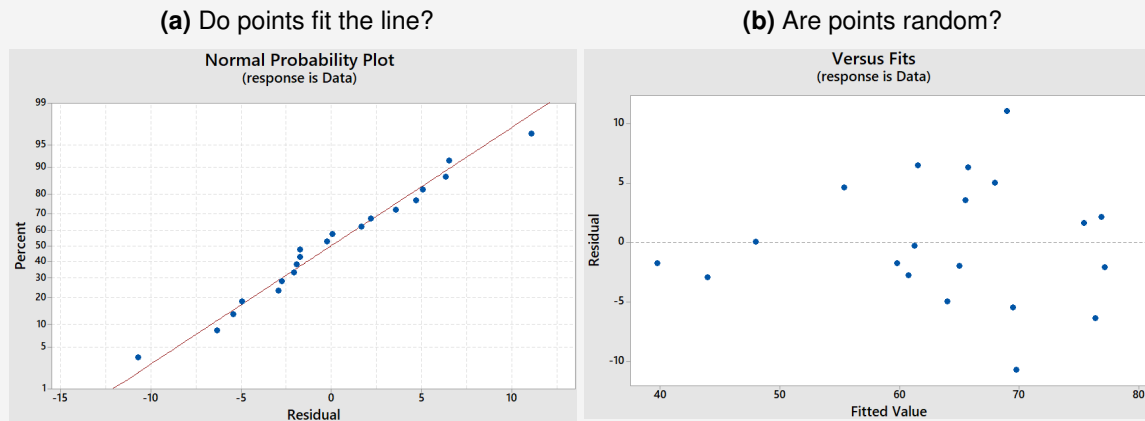
The following linear regression model resulted from Minitab's two-way ANOVA test results.

$$\begin{aligned}
 OR_{time} = & 77.10 - 15.60Location_3N - 11.40Location_3S \\
 & - 7.40Location_4E + 0.0Location_4W + 0.0Physician_PhysicianA \\
 & - 0.25Physician_PhysicianB - 1.75Physician_PhysicianC \\
 & - 0.75Physician_PhysicianD - 21.75Physician_PhysicianE
 \end{aligned}$$

We can use the coefficients of this regression model to infer how much effect different variables have on the OR time. Minitab indicates that only the coefficients of *Location_3N*, *Location_3S*, and *Physician_PhysicianE* are statistically significant. Accordingly, we can say that for each additional unit of procedure performed in the OR located at 3N, OR time decreases by 15.60 minutes on average when all other variables are held constant. Likewise, for each additional unit of procedure performed in the OR located at 3S, OR time decreases by 11.40 minutes on average, and for each additional unit of procedure performed by PhysicianE, OR time decreases by 21.75 minutes on average. In both cases, all other variables are held constant.

Residual plots

The residual plots of our regression model are shown in Figure 7.11. Subfigure 7.11a indicates that the residuals points fall alongside the fitted line, thus suggesting the normal distribution of the residuals. Moreover, the plot of the residuals versus the fitted values in Subfigure 7.11b seems to portray random behaviors, another confirmation of the normality of the residuals. Although, we do notice a mild resemblance to a funnel-like shape, which may be an indication of unequal variances. So, we cautiously conclude that our model is a good fit.

Figure 7.11: Residual plots of the OR_{time} regression model

7.8.3 Repeatability and Reproducibility (R&R) Analysis

One common application of the ANOVA test during quality improvement projects involves gauge analysis. Gauge analysis, also commonly referred to as **Gauge R&R** or **Gage R&R** study, is an essential activity of quality improvement that we typically undertake in the *Measure* and *Analyze* phases of DMAIC.

R&R

The main goal with R&R is to ensure that we do not introduce extra variability in quality measures due to inconsistent data collection methods. Our task here is to calibrate our gauge so that our measurements are repeatable and reproducible [43].

Definition 7.3 (Repeatability & Reproducibility)

1. **Repeatability** implies that measurements of the same unit taken under the **same** conditions (e.g., same operator) are consistent.
2. **Reproducibility** implies that measurements of the same unit taken under **different** conditions (e.g., different operators) are consistent.

The ANOVA model that applies to the gauge analysis includes the total sum of squares structured as follows [43]:

$$SS_{total} = SS_{parts} + SS_{operators} + SS_{parts \times operators} + SS_{error} \quad (7.75)$$

where *parts* are the units being measured and *operators* are individuals taking measurements. The subscript of $parts \times operators$ indicates the interaction between *parts* and *operators*. The component of SS_{total} that is due to the variation in parts is SS_{parts} , and the component indicating variation in the gauge is given by [43]:

$$SS_{gauge} = SS_{operators} + SS_{parts \times operators} + SS_{error} \quad (7.76)$$

The sum of $SS_{operators} + SS_{parts \times operators}$ in Equation 7.76 captures variability due to the lack of **reproducibility** of the measurements. The SS_{error} term represents variability due to the lack of **repeatability** of the measurements. It follows that the variability in the gauge can be expressed this way:

$$SS_{gauge} = SS_{reproducibility} + SS_{repeatability} \quad (7.77)$$

In variance units, the variability of the gauge looks as follows:

$$\sigma_{gauge}^2 = \sigma_{reproducibility}^2 + \sigma_{repeatability}^2 \quad (7.78)$$

The variance model of the total variance follows:

$$\sigma_{total}^2 = \sigma_{gauge}^2 + \sigma_{parts}^2 \quad (7.79)$$

During an R&R study, we attempt to calibrate our gauge to minimize σ_{gauge}^2 , which in turn reduces the overall variability σ_{total}^2 .

The C_p ratio

We have previously discussed the C_p ratio in Module 4. Besides measuring the process capability, this ratio can also be utilized to evaluate the acceptability of the *gauge* instrument. When inverted, the C_p ratio looks this way:

$$P/T = \frac{6\sigma_{gauge}}{USL - LSL} \quad (7.80)$$

where P/T stands for the **precision-to-tolerance** ratio and σ_{gauge} is obtained by taking the square root of σ_{gauge}^2 in Equation 7.78. In typical applications, $P/T \leq 0.1$ implies a well-calibrated gauge instrument [43].

How-To 7.12 (Gage R&R Study in Minitab 18)

Minitab has the following three options for conducting the Gage R&R analysis:

1. **Crossed:** this option is used when every operator can measure every part. In Minitab 18, go to: *Stat > Quality Tools > Gage Study > Gage R&R Study (Crossed)*.
2. **Nested:** this option is used when each operator is not able to measure all parts. In Minitab 18, go to: *Stat > Quality Tools > Gage Study > Gage R&R Study (Nested)*.
3. **Expanded:** this option is used to assess the influence of more factors, beyond parts and operators. In Minitab 18, go to: *Stat > Quality Tools > Gage Study > Gage R&R Study (Expanded)*.

Example 7.4 (R&R Study at Metropolis Hospital)

A quality manager at Metropolis Hospital would like to conduct a Gage R&R study to assess the consistency in the measurements of the clinical test X. For this study, twelve random patients and three random nurses volunteered to participate. Each nurse took the measurements three times on each patient. The results are presented in Table 7.7.

Table 7.7: Measurements of the clinical text X at Metropolis Hospital

<i>Patient</i>	<i>Nurse1</i>			<i>Nurse2</i>			<i>Nurse3</i>		
#	Test1	Test2	Test3	Test1	Test2	Test3	Test1	Test2	Test3
1	63	70	62	65	69	80	83	70	94
2	91	65	53	89	81	62	85	78	72
3	52	100	88	90	76	63	93	94	72
4	79	54	65	91	86	89	88	78	72
5	98	55	90	68	70	81	75	99	93
6	96	60	89	61	74	76	95	81	76
7	60	98	89	74	91	96	95	90	76
8	57	60	71	96	77	88	72	76	72
9	94	50	93	74	75	73	100	83	76
10	64	77	50	72	69	66	94	94	100
11	63	61	56	60	95	68	98	74	100
12	51	65	91	64	92	65	96	73	76

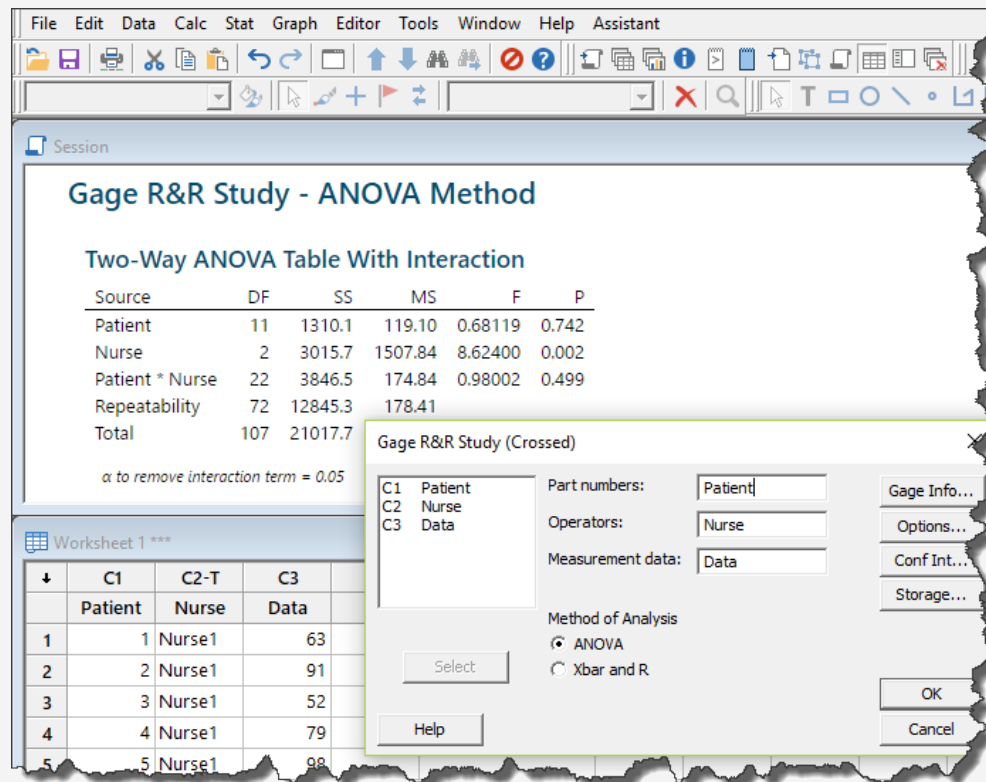
The manager rearranged the data, as illustrated in Table 7.8, and used Minitab 18 to conduct a crossed R&R analysis.

Table 7.8: Rearranged data for a crossed R&R analysis in Minitab 18

<i>Patient</i>	<i>Nurse</i>	<i>Data</i>
1	Nurse1	63
2	Nurse1	91
3	Nurse1	52
4	Nurse1	79
5	Nurse1	98
6	Nurse1	96
7	Nurse1	60
8	Nurse1	57
9	Nurse1	94
10	Nurse1	64
11	Nurse1	63
12	Nurse1	51
1	Nurse1	70
2	Nurse1	65
...
10	Nurse3	100
11	Nurse3	100
12	Nurse3	76

To generate ANOVA results, click on *Stat > Quality Tools > Gage Study > Gage R&R Study (Crossed)*. In the pop-up box, fill in Part numbers (Patient), Operators (Nurse), and Measurements (Data), as illustrated in Figure 7.12. Make sure the ANOVA radio button is selected.

Figure 7.12: Options for a Gage R&R study in Minitab 18



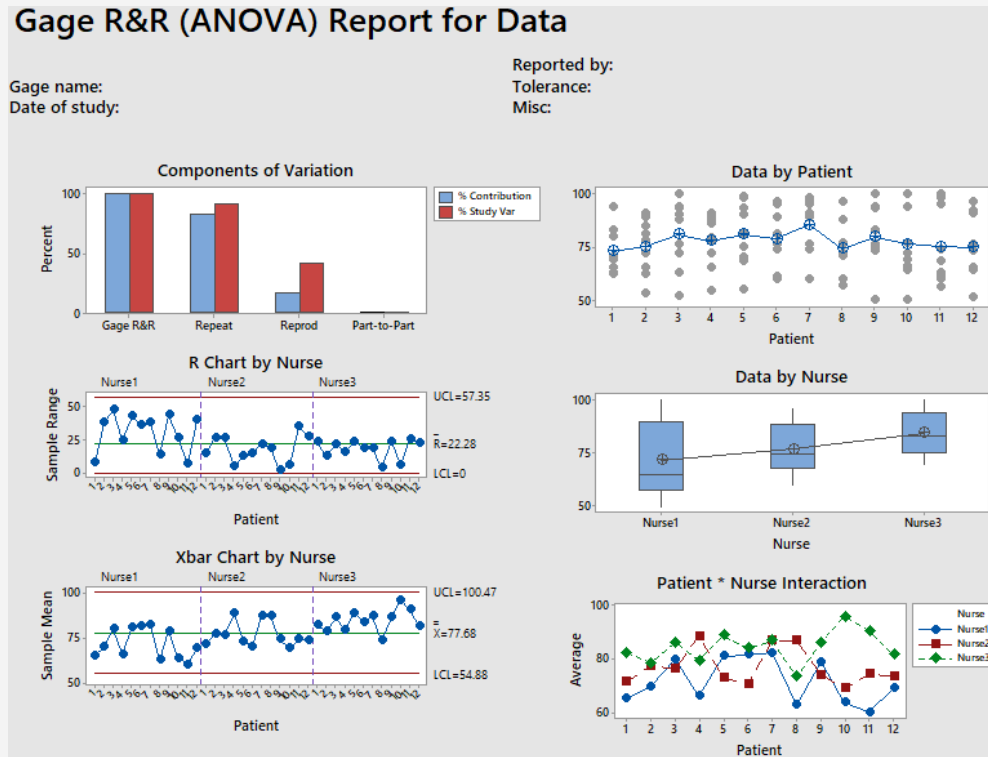
Minitab numerical results are presented in Table 7.9. Graphical results are depicted in Figure 7.13. The ANOVA results in Table 7.9 indicate that the factor of **Nurse** is statistically significant, given the $p\text{-value} < 0.05$. The interaction between patient and nurse (*Patient*Nurse*) is not statistically significant. The results of the **Variance Components** suggest that *repeatability* is the primary source of variability and contributed about 83% of the variance in the measurements. In other words, the nurses are unable to repeat their measurements. We arrive at the same conclusion using the standard deviation units in the **Gage Evaluation** results. Furthermore, the **Component of Variation** bar chart in Figure 7.13 suggests that almost 100% of the variation is due to Gage R&R, with repeatability contributing most of the variability. The **Patient*Nurse Interaction** chart suggests that for any given patient, there is noticeable inconsistency in the measurements (especially for patients 10 and 11), meaning that reproducibility is also a non-trivial factor in this study. The **R Chart by Nurse** captures the process variability by the nurse. The results suggest that the measurements of Nurse1 are the most inconsistent. Also, see how the interquartile range of Nurse1 is the biggest in the **Data by Nurse** box plots. The

Xbar Chart by Nurse shows a peculiar upward trend of the average of measurements. We expected all measurements to behave randomly around $\bar{\bar{X}}$, but it seems that the measurements of Nurse1 are mostly below average, whereas the measurements of Nurse3 are, for the most part, above average. So, what can we conclude? **We can conclude that the test instrument should be calibrated, and nurses retrained on taking and recording measurements of the clinical test X. This solution should be iterated until the test measurements are all reproducible and repeatable.**

Table 7.9: Minitab Gage R&R ANOVA Results

Two-Way ANOVA Table With Interaction						Variance Components		
Source	DF	SS	MS	F	P	Source	VarComp	%Contribution (of VarComp)
Patient	11	1310.1	119.1	0.6812	0.74	Total	Gage	214.525
Nurse	2	3015.7	1507.8	8.624	0	R&R		100
Patient * Nurse	22	3846.5	174.84	0.98	0.5	Repeatability	177.573	82.78
Repeatability	72	12845	178.41			Reproducibility	36.952	17.22
Total	107	21018				Nurse	36.952	17.22
α to remove interaction term = 0.05						Part-To-Part	0	0
						Total Variation	214.525	100

Two-Way ANOVA Table Without Interaction						Gage Evaluation			
Source	DF	SS	MS	F	P	Source	StdDev (SD)	Study Var (6 × SD)	%Study Var (%SV)
Patient	11	1310.1	119.1	0.6707	0.76	Total	Gage	14.6467	87.88
Nurse	2	3015.7	1507.8	8.4914	0	R&R			100
Repeatability	94	16692	177.57			Repeatability	13.3257	79.9539	90.98
Total	107	21018				Reproducibility	6.0788	36.4729	41.5
						Nurse	6.0788	36.4729	41.5
						Part-To-Part	0	0	0
						Total Variation	14.6467	87.88	100

Figure 7.13: Minitab Gage R&R Report

7.9 Chi-square test

A chi-square test is a statistical technique that we use to validate the goodness-of-fit of the hypothesized probability distribution of a given data set [57]. Within the context of quality improvement, we can apply the chi-square test to compare the current process performance to the expected standard.

7.9.1 One variable

To apply the chi-square test to compare the observed quality and the standard, we state our hypothesis as follows:

H_0 : there is no difference between the observed quality measures and the standard.

H_1 : there is a difference between the observed quality measures and the standard.

We obtain the chi-square test statistic χ_0^2 as follows:

$$\chi_0^2 = \sum \left[\frac{(f_0 - f_e)^2}{f_e} \right] \quad (7.81)$$

Here, f_0 is a vector of the observed quality measures and f_e is a vector of the expected quality measures. The test critical value is given by $\chi^2_{\alpha,df}$, where α is the significance level, and df indicates the degrees of freedom. We determine the df as follows:

$$df = (\text{numbers of columns} - 1)(\text{numbers of rows} - 1) \quad (7.82)$$

If $\chi^2_0 > \chi^2_{\alpha,df}$, we reject the null hypothesis H_0 . Alternatively, we can reject H_0 , if the p-value is less than the significance level α .

7.9.2 Two variables

When our process data consist of two distinct and nominally-scaled variables, we can apply the chi-square test to assess the significance of the relationship between these variables. In this case, it is easier to organize our data into a **contingency table**, with one variable in the rows and the other in the columns. We state our hypothesis this way:

H_0 : there is no relationship between two nominally-scaled variables in the contingency table.

H_1 : there is a relationship between two nominally-scaled variables in the contingency table.

Each cell in the contingency table represents the observed value or f_0 . We obtain the cell's expected value f_e as follows:

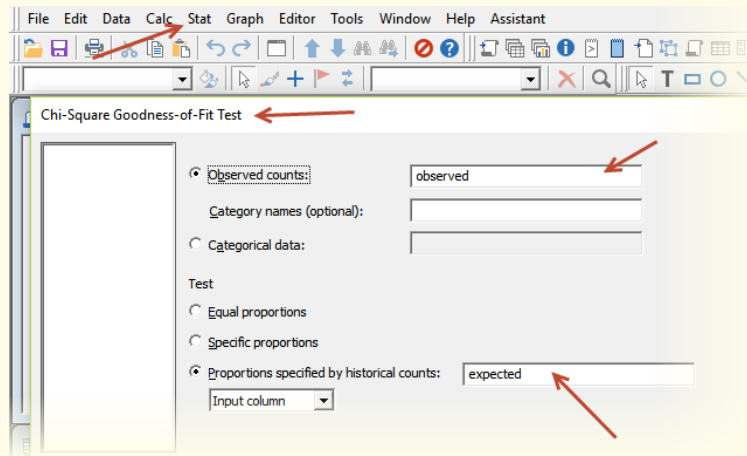
$$f_e = \frac{(\text{Row total})(\text{Column total})}{(\text{Grand total})} \quad (7.83)$$

The test statistic and degrees of freedom follow from Equations 7.81 and 7.82, respectively. The decision rule remains as before. We reject H_0 when $\chi^2_0 > \chi^2_{\alpha,df}$ or when $p\text{-value} < \alpha$.

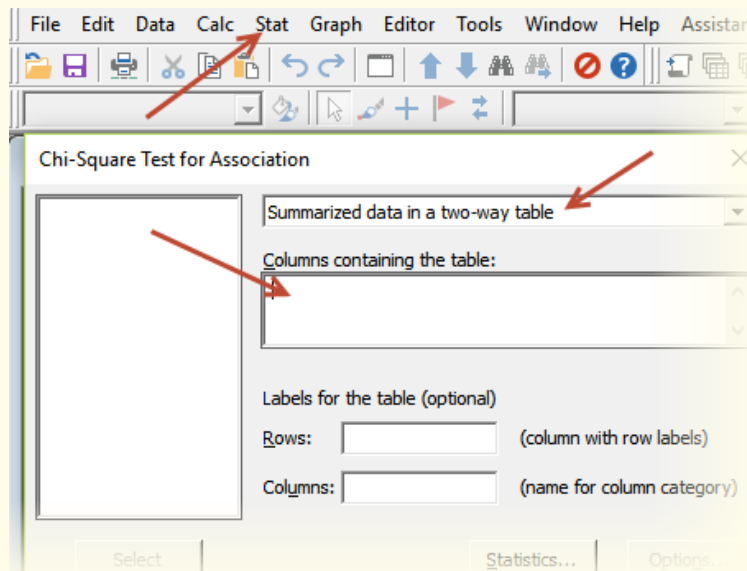
How-To 7.13 (Chi-square test in Minitab 18)

1. To run the chi-square test for a single variable, click on *Stat > Tables > Chi-Square Goodness-of-Fit Test > input your data as shown in Figure 7.14 > OK*.



Figure 7.14: Options for the chi-square test for a single variable in Minitab 18

2. To run the chi-square test for a contingency table, click on *Stat > Tables > Chi-Square Test for Association*> In the drop-down menu, select *Summarized data in a two-way table* > input your data as shown in Figure 7.15 > OK.

Figure 7.15: Options for the chi-square test for a contingency table in Minitab 18**How-To 7.14 (Chi-square test in Excel 2013)**

1. To find the critical value $\chi^2_{\alpha, df}$, use this function: $=CHISQ.INV.RT(\alpha, df)$
2. To find the p-value of the chi-square test, use this function: $=CHISQ.DIST.RT(\chi_0, df)$

Example 7.5 (Chi-square test on a single variable)

Consider the data in Table 7.10, where the inpatient coding productivity rate is recorded by *Coder* and *Expected* standard.

Table 7.10: Inpatient coding productivity rates

Coder	Observed	Expected
A	2.0	2.4
B	2.5	2.4
C	2.9	2.4
D	1.7	2.4
E	1.2	2.4
F	4.1	2.4

Your interest is to test whether there is a statistical difference between the observed and the expected productivity rate. At the significance level of $\alpha = 0.05$ and $df = 5$, you found the critical value in Excel using $CHISQ.INV.RT(0.05,5) = 11.07$. To find the chi-square test statistic χ_0^2 , you set up your Excel, as shown in Table 7.11.

Table 7.11: The Excel setup for the chi-square test of the data in Table 7.10

Coder	Observed	Expected	(Observed - Expected) ² /Expected
A	2.0	2.4	0.0667
B	2.5	2.4	0.0042
C	2.9	2.4	0.1042
D	1.7	2.4	0.2042
E	1.2	2.4	0.6
F	4.1	2.4	1.2042

$$\chi_0^2 =$$

$$\sum column = 2.183$$

Given $\chi_0^2 < \chi_{\alpha, df}^2$ you failed to reject the null hypothesis that there is no difference between the observed and the expected coding productivity rates. Correspondingly, the p-value is greater than α , as follows: $= CHISQ.DIST.RT(2.183, 5) = 0.823 > 0.05$. Minitab 18 produces similar results as shown in Figure 7.16.

Figure 7.16: The chi-square test in Minitab 18 based on the data in Table 7.10**Chi-Square Goodness-of-Fit Test for Observed Counts in ... ble: Actual****Observed and Expected Counts**

Category	Observed	Historical Counts	Test Proportion	Expected	Contribution to Chi-Square
1	2.0	2.4	0.166667	2.4	0.06667
2	2.5	2.4	0.166667	2.4	0.00417
3	2.9	2.4	0.166667	2.4	0.10417
4	1.7	2.4	0.166667	2.4	0.20417
5	1.2	2.4	0.166667	2.4	0.60000
6	4.1	2.4	0.166667	2.4	1.20417

6 (100.00%) of the expected counts are less than 5.

Chi-Square Test

N	DF	Chi-Sq	P-Value
14.4	5	2.18333	0.823

Example 7.6 (Chi-square test for a contingency table)

You audited 100 medical records and recorded documentation issues by the record type, as shown in Table 7.12. You classified each record as an inpatient, outpatient, or ancillary. You denoted issues by letters A (mixed messages), B (misuse of copy and paste), C (incomplete information), and D (misplaced documentation) [15]. Your interest is to test whether there is a relationship between the record type and documentation issues. To do that, you perform a chi-square test.

Table 7.12: Documentation issues by the record type

	A	B	C	D
Inpatient	17	21	46	12
Outpatient	8	1	24	3
Ancillary	34	15	47	20

The chi-square test results from Minitab are displayed in Table 7.13.

Table 7.13: The chi-square test results from Minitab 18 based on data in Table 7.12

	Chi-Square	DF	P-Value
Pearson	15.827	6	0.015
Likelihood Ratio	17.138	6	0.009

Given the p-values in this case, you reject the null hypothesis. In other words, a relationship likely exists between the record type and documentation

issues. To reproduce the Pearson p-value in Excel, proceed as follows: $=CHISQ.DIST.RT(15.827,6) = 0.015$.

7.10 League tables, funnel charts, and ANOM

League tables, funnel charts, and the analysis of means (ANOM) are among many tools that we can utilize to benchmark quality from different health care organizations.

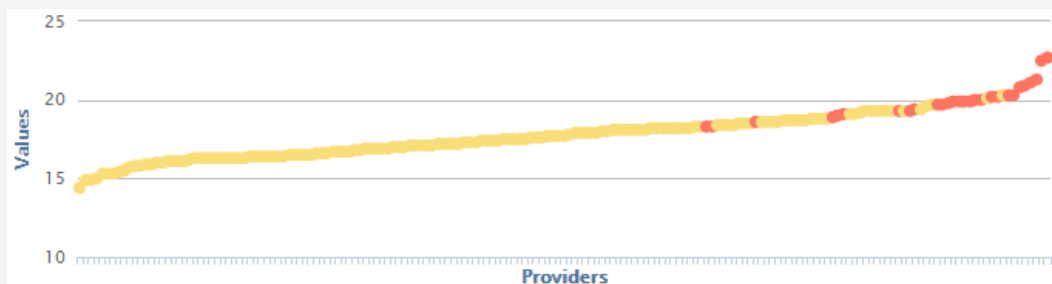
7.10.1 A league table

A league table is one of the classical methods for benchmarking quality in health care. With this technique, organizations are ranked and ordered from the highest to the lowest performer [56]. The main criticism levied against this method is the failure to account for the sample size and the confidence interval [56, 61]. Various other methods for ranking health care organizations exist, such as quality report cards published by the federal [41] and state governments [48]. Private organizations such as the Leapfrog Group [26] also do publish various rankings of health care organizations.

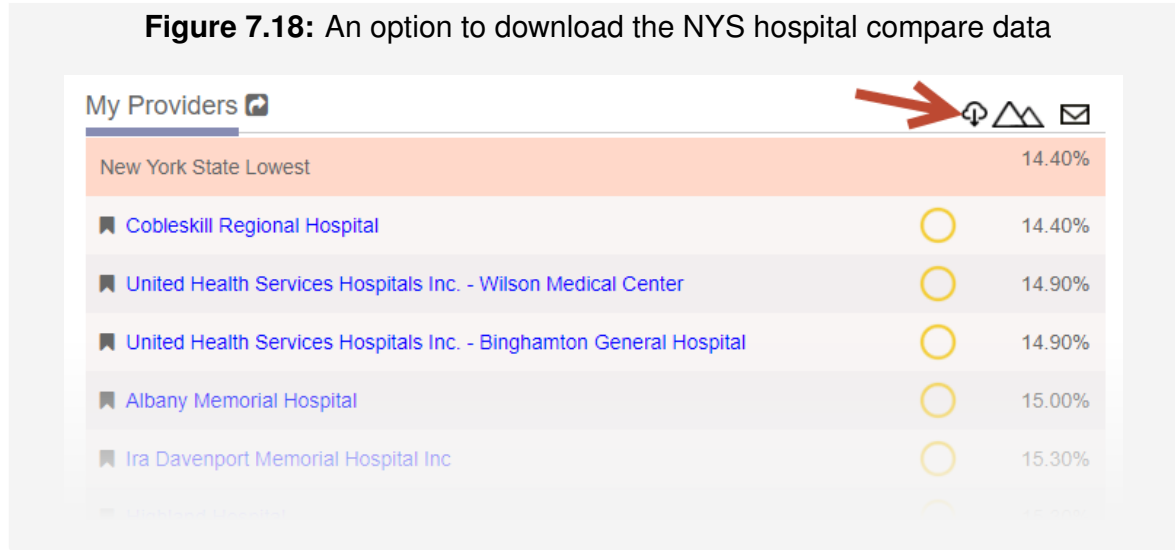
Example 7.7 (Public report cards)

To view an example of public report cards concerning readmission rates of pneumonia patients in New York State (NYS), visit <https://profiles.health.ny.gov/hospital>. Next, select Hospitals > Compare Hospitals. Under *Select Domain Type*, choose *Readmission Within 30 Days*. Under *Select Measure*, choose *Pneumonia Patients Readmission Within 30 Days*. A chart that looks like Figure 7.17 will appear.

Figure 7.17: Readmission rates of pneumonia patients, within 30 days, NYS hospitals. The x-axis represents hospitals, and the y-axis characterizes the percent of readmission. Source: <https://profiles.health.ny.gov/measures/view/10083>, Report Period: July 1, 2014, to June 30, 2017



To see how the rates and the ranking scores were calculated, click on the legend link, which may take you on this web page: <https://profiles.health.ny.gov/hospital/pages/methodology>. To download the data, click on the cloud symbol with a downward arrow (see the snapshot in Figure 7.18).

Figure 7.18: An option to download the NYS hospital compare data

7.10.2 Funnel charts

Like control charts, funnel charts are constructed using control limits and a center line. A typical funnel chart tends to look like a p chart and is structured as follows [55]:

$$UCL = p + L\sqrt{\frac{p(1-p)}{n_i}} \quad (7.84)$$

$$CL = p \quad (7.85)$$

$$LCL = p - L\sqrt{\frac{p(1-p)}{n_i}} \quad (7.86)$$

Here, p is the target, L is the number of standard deviations away from the centerline, commonly chosen to create a 95% or 99.8% confidence interval [61]. The notation of n_i signifies the sample size of organization i , for $i : 1, 2, \dots, k$, where k is the total number of organizations being compared. The acronyms UCL , CL , and LCL , stand for *upper control limit*, *centerline*, and *lower control limit* respectively. If a point falls outside of a particular limit, we conclude that the deviation of the corresponding sample number, away from the target or the grand mean, is statistically significant. The only difference between a funnel chart and a p chart is that in the former, we have to sort samples in ascending order, a practice that we do not adhere to in the latter. We will discuss more about a p chart in Module 9.

7.10.3 ANOM

ANOM, not to be confused with ANOVA, is a statistical technique that we can employ to benchmark quality from several health care organizations. The main difference between ANOVA and ANOM is that the former compares the equality of means, whereas the latter compares each mean to the overall mean. Just like in funnel charts, the formulas for

ANOM charts often look like those of a p-chart [55]. In Minitab 18, the formulas for the binomial ANOM chart look like this [1]:

$$UDL = p + L \sqrt{\frac{p(1-p)}{n}} * \sqrt{\frac{k-1}{k}} \quad (7.87)$$

$$LDL = p - L \sqrt{\frac{p(1-p)}{n}} * \sqrt{\frac{k-1}{k}} \quad (7.88)$$

Here, *UDL* stands for *upper decision limit* and *LDL* means *lower decision limit*. The notation of *n* symbolizes a constant sample size, and *k* represents the number of organizations in the study. As before, *L* represents the number of standard deviations away from the centerline *p*. Minitab 18 determines *L* as follows:

$$L = \Phi^{-1}(1 - \alpha/(2 * k)) \quad (7.89)$$

where * symbol characterizes multiplication and $\Phi(\cdot)$ is the cumulative density function of a standard normal. As before, α is the given significance level. As in funnel charts, if a point falls outside of a particular limit, we conclude that the corresponding sample is statistically different from the target or the grand mean.

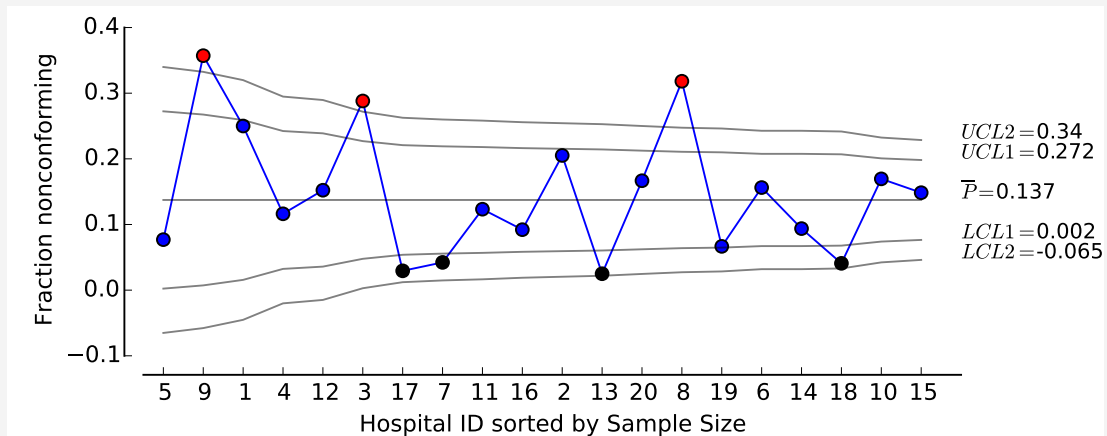
How-To 7.15 (Funnel charts in Excel 2013)

Excel does not have a built-in option for creating ANOM and funnel charts, but we can manually program the respective formulas as we demonstrate in Example 7.8.



Example 7.8 (A funnel chart in Excel 2013)

The league table given in Table 7.14 depicts the ranking of 20 hospitals in Central City, regarding the rates of pneumonia patients readmitted within 30 days of discharge. The lower the rate, the better. The average readmission rate is 13.7%. Hospitals that are above this average are marked in red. Table 7.15 demonstrates the setup of an Excel spreadsheet for creating a funnel chart based on the data in Table 7.14. In this setup, *ID* denotes the *hospital ID* column, *n_i* symbolizes the column of *Pneumonia Discharges* or sample size, *x_i* represents the column of *Readmit with 30 days*, and *x_i/n_i* represents to the *Rate* column. We calculated the average rate using $p = \sum_{i=1}^k x_i/n_i$, where *k* is the total number of hospitals. Before plotting the funnel chart, we sorted Table 7.15 by the sample size *n_i*. We applied Equations 7.84 - 7.86 to create inner control limits (*LCL₁* and *UCL₁* using *L* = 2) and outer control limits (*LCL₂* and *UCL₂* using *L* = 3). For example, we obtained the value of *LCL₁* in cell E2 using $= I2 - 2 * SQRT(I2 * (1 - I2)/B2)$ and the value of *UCL₂* in cell H2 using $= I2 + 3 * SQRT(I2 * (1 - I2)/B2)$. From Table 7.15, we created a funnel chart by inserting line charts of columns *D* to *I*, as illustrated in Figure 7.19. From this chart, we see that hospitals with IDs 3, 8, and 9 fail *UCL₂*. Also, we see that hospitals with IDs 7, 13, 17, and 18 fail *LCL₁*.

Figure 7.19: A funnel chart based on the data in Table 7.14**Table 7.14:** Pneumonia patients readmitted within 30 days of discharge in Central City, 2018

Hospital ID	Pneumonia Discharges	Readmit with 30 days	Rate	Rank
13	80	2	2.5%	1
17	68	2	2.9%	2
18	98	4	4.1%	3
7	71	3	4.2%	4
19	90	6	6.7%	5
5	26	2	7.7%	6
16	76	7	9.2%	7
14	96	9	9.4%	8
4	43	5	11.6%	9
11	73	9	12.3%	10
15	128	19	14.8%	11
12	46	7	15.2%	12
6	96	15	15.6%	13
20	84	14	16.7%	14
10	118	20	16.9%	15
2	78	16	20.5%	16
1	32	8	25.0%	18
3	59	17	28.8%	17
8	88	28	31.8%	19
9	28	10	35.7%	20

Table 7.15: A setup of an Excel spreadsheet for creating a funnel chart based on the data in Table 7.14

	A	B	C	D	E	F	G	H	I
1	<i>ID</i>	<i>n_i</i>	<i>x_i</i>	<i>x_i/n_i</i>	<i>LCL₁</i>	<i>UCL₁</i>	<i>LCL₂</i>	<i>UCL₂</i>	<i>p</i>
2	5	26	2	7.7%	0.2%	27.2%	-6.5%	34.0%	13.7%
3	9	28	10	35.7%	0.7%	26.7%	-5.8%	33.2%	13.7%
4	1	32	8	25.0%	1.6%	25.9%	-4.5%	32.0%	13.7%
5	4	43	5	11.6%	3.2%	24.2%	-2.0%	29.5%	13.7%
6	12	46	7	15.2%	3.6%	23.9%	-1.5%	29.0%	13.7%
7	3	59	17	28.8%	4.8%	22.7%	0.3%	27.2%	13.7%
8	17	68	2	2.9%	5.4%	22.1%	1.2%	26.3%	13.7%
9	7	71	3	4.2%	5.6%	21.9%	1.5%	26.0%	13.7%
10	11	73	9	12.3%	5.7%	21.8%	1.6%	25.8%	13.7%
11	16	76	7	9.2%	5.8%	21.6%	1.9%	25.6%	13.7%
12	2	78	16	20.5%	5.9%	21.5%	2.0%	25.4%	13.7%
13	13	80	2	2.5%	6.0%	21.4%	2.2%	25.3%	13.7%
14	20	84	14	16.7%	6.2%	21.2%	2.5%	25.0%	13.7%
15	8	88	28	31.8%	6.4%	21.1%	2.7%	24.7%	13.7%
16	19	90	6	6.7%	6.5%	21.0%	2.8%	24.6%	13.7%
17	14	96	9	9.4%	6.7%	20.8%	3.2%	24.3%	13.7%
18	6	96	15	15.6%	6.7%	20.8%	3.2%	24.3%	13.7%
19	18	98	4	4.1%	6.8%	20.7%	3.3%	24.2%	13.7%
20	10	118	20	16.9%	7.4%	20.1%	4.2%	23.2%	13.7%
21	15	128	19	14.8%	7.6%	19.8%	4.6%	22.9%	13.7%

In contrast to the hospital ranking in Table 7.14 where ten hospitals may be considered deficient since they are above the average, the funnel chart indicates that only a few hospitals fail the control limits at 99.73% confidence interval. In typical applications, the negative part of the funnel limits is set to zero; we plotted it here only to demonstrate the funnel shape from which the name *funnel chart* is derived.

How-To 7.16 (ANOM chart in Minitab 18)

Click on *Stat > ANOVA > Analysis of Means*. Select your data for the *Response* field. For binomial data, make sure you select *Binomial*. Minitab assumes a constant sample size. Input the appropriate sample size in the designated area. See an illustration in Figure 7.20.



Figure 7.20: ANOM Options in Minitab 18

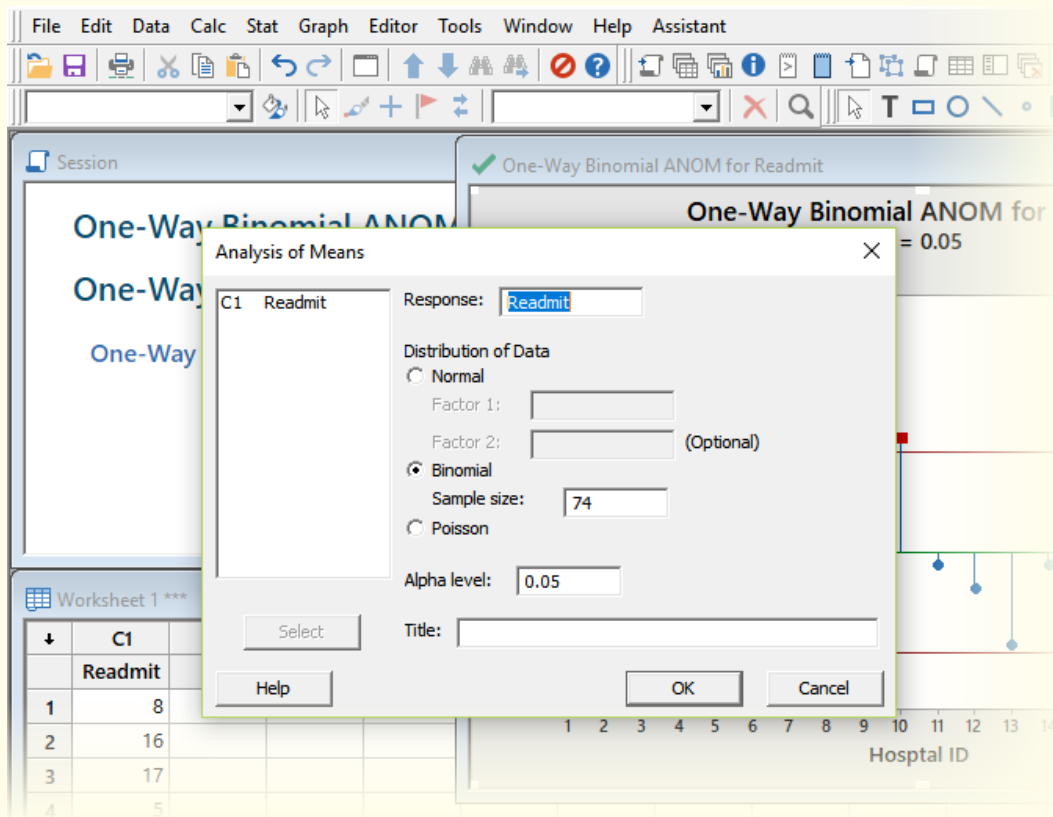
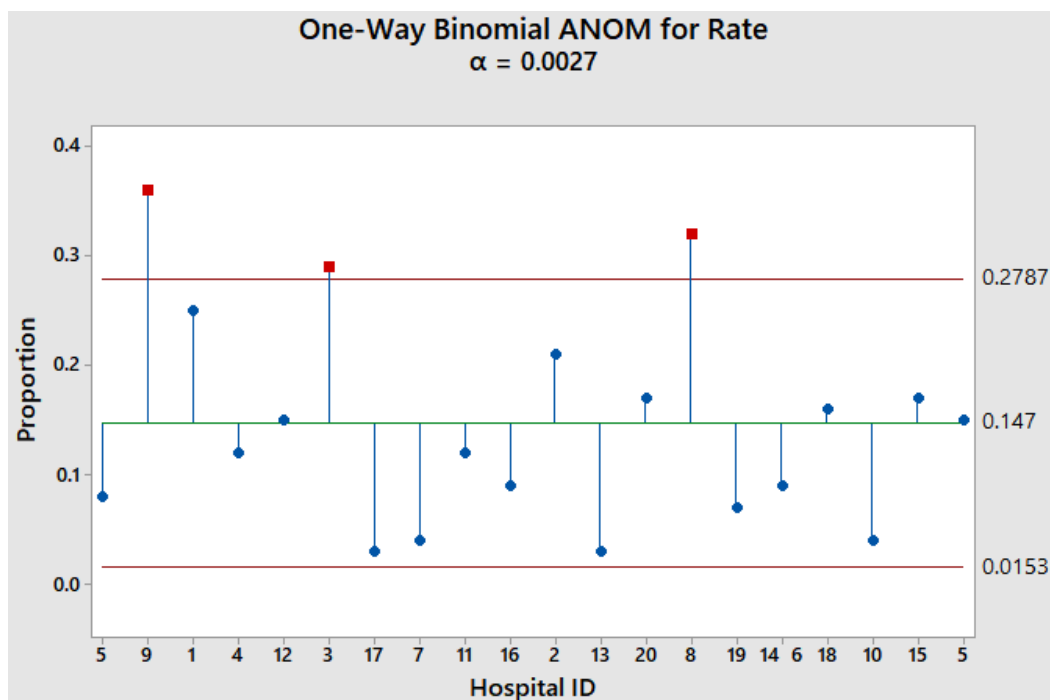


Figure 7.21: ANOM chart based on the readmission data in Table 7.14



Example 7.9 (ANOM in Minitab 18)

For demonstration purposes, we used Minitab 18 to approximate an ANOM chart related to the data in Table 7.14. Since Minitab assumes a constant sample size for all subgroups, we chose the size of 100 and transformed all rates to relate to this constant. For example, we changed the rate of 7.7%, in cell D2 of Table 7.15, to $ROUND(7.7\% * 100, 0) = 8$. In Minitab, we input 100 as the binomial constant sample size. Figure 7.21 illustrates the resulting ANOM chart at $\alpha = 0.0027$. From Figure 7.21, we observe that three points fall outside of the control limits. In both the funnel and ANOM charts, the out-of-control points correspond to hospitals with ID 9 ($n_i = 28$), ID 3 ($n_i = 59$), and ID 8 ($n_i = 88$). Assuming comparable processes, we conclude that readmission rates in these hospitals are statistically different from the overall mean rate of 14.7%, at the 99.73% confidence level.

7.11 Check Your Understanding

1. A false alarm in the process is equivalent to:
 - (a) type I error
 - (b) type II error
2. When comparing the means in paired observations, we typically apply which test?
 - (a) t-test
 - (b) Z-test
 - (c) F-test
 - (d) Chi-square-test
3. The ratio between the two sample variances is likely to follow which distribution?
 - (a) normal distribution
 - (b) t-distribution
 - (c) F-distribution
 - (d) Chi-square distribution
4. Our null hypothesis says that $H_0 : \sigma_1^2 = \sigma_2^2$. What statistic test are we likely to run?
 - (a) F-test
 - (b) t-Test
 - (c) Z-test
 - (d) Chi-square-test
5. When performing an ANOVA test, as the degrees of freedom in the error decrease, we are likely to conclude that:
 - (a) There is no difference in the population means
 - (b) There is a difference in the population means
6. When performing a chi-square test from a contingency table, how do you obtain the f_e value of a cell?
7. What statistical test can you apply to investigate the null hypothesis that there is no difference between the observed and the expected quality performance?
8. Using Excel, verify the Chi-Square test results in Example 7.6. What was your critical value?
9. Which of the following tools can you apply to rank health care organizations from best to worst:

- (a) League tables
 - (b) Funnel charts
 - (c) ANOM charts
 - (d) ANOVA test
10. What can you conclude when a point falls outside of a UDL of an ANOM chart?
11. How does Minitab handle sample sizes for binomial data when creating ANOM charts?
12. In what way does a funnel chart differ from a p-control chart?
13. At the confidence interval of 99.73%, what is the LDL of an ANOM chart with $p = 0.10$, the constant sample size is 100, and there are 10 health care organizations in the study?
- (a) 0.0123
 - (b) 0.0146
 - (c) 0.0253
 - (d) 0.0398
14. During an R&R study, we typically employ the following statistical test:
- (a) ANOVA
 - (b) ANOM
 - (c) Chi-square
 - (d) Z
15. The variance of the gauge is estimated from which components of the total variation (circle all that apply):
- (a) Operators
 - (b) Operators*parts
 - (c) Error
 - (d) Parts
16. When an operator can measure every part, we select the following R&R option in Minitab:
- (a) Nested
 - (b) Crossed
 - (c) Expanded
 - (d) Comprehensive

17. Your two-way ANOVA test results from Minitab indicates an $R^2_{adjusted}$ of the regression model to be 13%. What does this percent signify about the statistical significance of the factors in the model?
18. In a two-way ANOVA test, if the degrees of freedom of the blocking variable were $b - 1$ and the degrees of freedom of the treatments were $a - 1$, what would be the degrees of freedom of SS_{error} ?
19. You audited 100 records from your EHR system and found several quality issues that you categorized, as shown in Table 7.16.

Table 7.16: The number of electronic records with data quality issues by category

Category	#issues
Completeness	30
Currency	13
Accuracy	41
Concordance	18

- (a) Based on Table 7.16, state your chi-square decision rule at $\alpha = 0.05$.
 - (b) From Table 7.16, if each category is expected to occur with probability $1/4$, what can you conclude about $H_0 : p = 1/4$ and $H_1 : p \neq 1/4$? Show your work.
20. You collected and categorized defects from a health informatics process, as shown in Table 7.17.

Table 7.17: Health informatics defects

Department\Defect	A	B	C	D	E
1	14	15	22	21	24
2	18	18	39	47	16
3	21	31	30	36	10
4	40	12	33	29	39
5	14	24	12	36	42

What can you conclude about the relationship between the department and the type of defects? First, state the null hypothesis, and then apply the appropriate decision rule to conclude about your hypothesis. You can use Excel or Minitab.

21. The quality manager at Metropolis Hospital has determined that nurse2 should not have taken part in the R&R study presented in Example 7.4. Remove nurse2 from

the study and rerun the R&R study in Minitab. Submit your Minitab file with the interpretation of your statistical and graphical results.

22. After further analysis, it was determined that hospitals with the IDs of 3, 8, and 9 were not in the same statistical state as the rest of the hospitals. Remove these hospitals from Table 7.14 and:
- (a) create a new funnel chart in Excel.
 - (b) approximate your results using the ANOM technique in Minitab.
 - (c) estimate the C_{pu} ratio of the entire process given $USL = 30\%$. What can you conclude?

Submit both your Excel and Minitab files with the interpretation of your results.

23. The results from an R&R study showed a p-value for the gauge to be 0.02 at the significance level of 0.05. All other p-values were above 0.05. This fact means that the major source of variability in the measurements is coming from:
- (a) Parts
 - (b) Operators
24. The following statistical test is used when conducting ANOVA:
- (a) F
 - (b) t
 - (c) Z
 - (d) χ^2
25. During an ANOVA test, you failed to reject H_0 when it was false. You likely committed which type of error:
- (a) Type I error
 - (b) Type II error

MODULE 8

Tools for improving the process

In this module, we review techniques for improving the process. The major statistical methods that we consider relate to the design of experiments (DOE). We also discuss Lean techniques to improve efficiency, and Poka-yoke measures to mitigate adverse events. Other discussion topics include the Swiss-cheese model and best practices for improving data quality.

Key concepts and tools: *Quality; Variability; Quality improvement; Process improvement; Data quality; Factorial design; DOE; Efficiency; Cycle time; Setup problems; Lean methods; ; Poka-yoke measures; 5S+1; Analysis of Variance (ANOVA); Control charts*

Major objectives

After studying this module, you will be able to:

- 1. Define key concepts and tools for improving the process*
- 2. Justify the need for designing experiments*
- 3. Design experiments*
- 4. Interpret ANOVA results from a DOE model*
- 5. Discuss typical setup problems in service*
- 6. Implement Lean solutions for efficiency*
- 7. Survey data quality improvement techniques*
- 8. Discuss the concept of the Swiss-cheese model*
- 9. Advocate for poka-yoke measures*
- 10. Manage quality improvement efforts*

8.1 Introduction

The *Improve* phase of DMAIC is probably the most challenging since we have to turn our analysis into actions to achieve the project goals that we introduced in the *Define* phase. To find feasible solutions, we may once again have to hold brainstorming sessions to generate ideas for improvement. For common efficiency problems, we could implement general *Lean* solutions to reduce the process cycle time. For solutions that require evaluating different combinations of the process inputs, we employ the design of experiments (DOE). Other process improvement concepts to consider include poka-yoke measures and the Swiss cheese model to mitigate adverse events. In health care informatics, we also must contemplate best practices and standards to improve data quality. In the end, we must reduce variability in the process and verify our success using control charts.

8.2 DOE

8.2.1 Background

DOE is a statistical technique that we employ during the *Improve* phase of DMAIC to attempt to reduce variability in the process. Two fundamental activities of DOE include the *experimentation* of different process inputs and the *comparison* of outputs using methods such as ANOVA. The goal is to select a statistically significant experiment that improves quality. The concept of DOE was first introduced by a statistician named Ronald Fisher in the 1920s, and this technique is widely applied in various other fields of study besides quality improvement.

8.2.2 The procedure for DOE

The procedure for designing experiments follows [43]:

1. **State the goal of running experiments:** this step is probably the most critical in DOE since unnecessary or incorrect investigations could imperil or slow down quality improvement efforts. So, our goal for DOE should be unambiguous and justifiable.
2. **Determine the key inputs variables (also known as *Factors*) and the primary output variable (also called *Response*):** in this step, we investigate the settings of all input variables that could potentially affect the output variable.
3. **Design experiments:** this phase, also referred to as the **factorial design**, involves determining the order of experiments, the number of replicates, the number of blocks, and the randomization techniques.
4. **Run experiments:** here, we run experiments and record the results according to our factorial design.

5. **Analyze the results:** in this step, we compare outputs using the ANOVA¹ techniques that we discussed in Module 7. Additionally, we use regression methods to evaluate the amount of variation in the response that is explained by the inputs.
6. **Decide on the course of action:** lastly, we choose statistically significant settings of the inputs (if any) to improve the process.

8.2.3 Factorial design

There are several approaches to designing experiments. The most common and probably the simplest is the 2^k design, where we only consider two settings of k factors. That is, for each factor, we investigate two possible levels that we designate as **low** (- or -1) or **high** (+ or 1). For example, if $k = 2$, we run $2^2 = 4$ experiments (or trials) to account for all possible combinations of the settings. Likewise, if $k = 3$, we run $2^3 = 8$ trials, and for $k = 4$ we run $2^4 = 16$ trials, and so on. It is apparent that as we increase the number of factors, the number of trials increases exponentially, a practical dilemma that we later discuss how to mitigate using a *fractional factorial design*.

Next, we provide details of 2^2 and 2^3 factorial designs.

A 2^2 factorial design

Let's assume that our process has two main factors, **A** and **B**, and that each factor has two levels, – and +. Table 8.1 presents the experiments to run for this design.

Table 8.1: Experiments in a 2^2 factorial design

Run#	Output	A	B
1	(1)	-	-
2	<i>a</i>	+	-
3	<i>b</i>	-	+
4	<i>ab</i>	+	+

The **Run#** column of Table 8.1 indicates the order of experiments. The **Output** column symbolizes experimental results by the run. We use letter *a* to denote the output when *A* is set to + and *B* is to –. We use letter *b* to denote the output when *B* is set to + and *A* is to –. Here, *ab* represents the output when both *A* and *B* are set to high (+), whereas (1) symbolizes the output when both factors are set to low (–). We construct a regression model from this factorial design as follows [43]:

$$Y = \beta_0 + \beta_1 A + \beta_2 B + \beta_{12} AB + \epsilon \quad (8.1)$$

where *A*, *B*, and *AB* are independent variables or **regressors**, *Y* is the dependent variable, and ϵ is the regression error. We obtain the intercept coefficient β_0 by taking the

¹ANOVA: Analysis of Variance

average of all runs. We determine other coefficient as follows:

$$\beta_1 = \frac{E_1}{2}; \beta_2 = \frac{E_2}{2}; \beta_{12} = \frac{E_{12}}{2} \quad (8.2)$$

where E_1 , E_2 , and E_{12} are the regression **effects** that we compute this way:

$$E_1 = \frac{a + ab - b - (1)}{2n} \quad (8.3)$$

$$E_2 = \frac{b + ab - a - (1)}{2n} \quad (8.4)$$

$$E_{12} = \frac{ab + (1) - a - b}{2n} \quad (8.5)$$

We use n to represent the number of **replicates** or simply the number of times we repeat the experiments. The numerators in Equations 8.3-8.5 are referred to as **contrasts**. Consequently, we can write any given effect E in terms of its contrast, as follows [43]:

$$E = \frac{\text{contrast}}{n2^{k-1}} \quad (8.6)$$

Furthermore, we can formulate the sum of squares (SS) of each factor this way:

$$SS = \frac{(\text{contrast})^2}{n2^k} \quad (8.7)$$

Following from the ANOVA discussion in Module 7, we can decompose the total sum of squares of our factorial design this way:

$$SS_{total} = SS_A + SS_B + SS_{AB} + SS_{error} \quad (8.8)$$

$$= \sum_{i=1}^{n_a} \sum_{j=1}^{n_b} \sum_{k=1}^n (y_{ijk} - \bar{y}_{...})^2 \quad (8.9)$$

Here, n_a is the number of settings of factor A , n_b is the number of settings of factor B , y_{ijk} is a particular value of a run in cell ij of replicate k , and $\bar{y}_{...}$ is the average of all runs. In a 2^2 design, $n_a = n_b = 2$. The degrees of freedom (df) of A , B , AB , and the *error* term are respectively given by $(n_a - 1)$, $(n_b - 1)$, $(n_a - 1)(n_b - 1)$, and $n_a n_b (n - 1)$. The total number of degrees of freedom equals $n_a n_b n - 1$ [43]. As before, we determine the mean square (MS) like this:

$$MS = \frac{SS}{df} \quad (8.10)$$

We obtain the test statistic (F_0) of any regressor as follows:

$$F_0 = \frac{MS_{regressor}}{MS_{error}} \quad (8.11)$$

As before, if F_0 is greater than the critical value of the F -distribution (at some given confidence level α), or when $p\text{-value} \leq \alpha$, we conclude that the related effect is statistically significant.

A 2^3 factorial design

In a 2^3 factorial design, we have three variables **A**, **B**, and **C** that we set according to Table 8.2.

Table 8.2: Experiments in a 2^3 factorial design

Run#	Output	A	B	C
1	(1)	-	-	-
2	<i>a</i>	+	-	-
3	<i>b</i>	-	+	-
4	<i>ab</i>	+	+	-
5	<i>c</i>	-	-	+
6	<i>ac</i>	+	-	+
7	<i>bc</i>	-	+	+
8	<i>abc</i>	+	+	+

We estimate the effects of this design as follows [43]:

$$E_1 = \frac{a + ab + ac + abc - b - c - bc - (1)}{4n} \quad (8.12)$$

$$E_2 = \frac{b + ab + bc + abc - a - c - ac - (1)}{4n} \quad (8.13)$$

$$E_3 = \frac{c + ac + bc + abc - a - b - ab - (1)}{4n} \quad (8.14)$$

$$E_{12} = \frac{ab + (1) + abc + c - b - a - bc - ac}{4n} \quad (8.15)$$

$$E_{13} = \frac{ac + (1) + abc + b - a - c - ab - bc}{4n} \quad (8.16)$$

$$E_{23} = \frac{bc + (1) + abc + a - b - c - ac}{4n} \quad (8.17)$$

$$E_{123} = \frac{abc - bc - ac + c - ab + b + a - (1)}{4n} \quad (8.18)$$

The corresponding regression equation looks this way:

$$Y = \beta_0 + \beta_1 A + \beta_2 B + \beta_3 C + \beta_{12} AB + \beta_{13} AC + \beta_{23} BC + \beta_{123} ABC + \epsilon \quad (8.19)$$

As earlier, we obtain the coefficient β_0 by taking the overall average of all runs. We determine other coefficients as follows: $\beta_1 = E_1/2$, $\beta_2 = E_2/2$, $\beta_3 = E_3/2$, $\beta_{12} = E_{12}/2$, $\beta_{13} = E_{13}/2$, $\beta_{23} = E_{23}/2$, and $\beta_{123} = E_{123}/2$. Like in the 2^2 design, we can use the contrasts to estimate the sum of squares and apply the F test to assess the statistical significance of each regressor.

2^k fractional factorial design

As the number of factors increases, the number of runs grows exponentially. A fractional factorial design can help mitigate this predicament. A general formulation of such design

is as follows:

$$2^{k-p} \quad (8.20)$$

where p is a constant that we choose to allow for the estimation of the main effects and some two-way interactions. For example, a 2^{k-1} fractional design implies that, instead of running all 2^k experiments, we only run half of them.

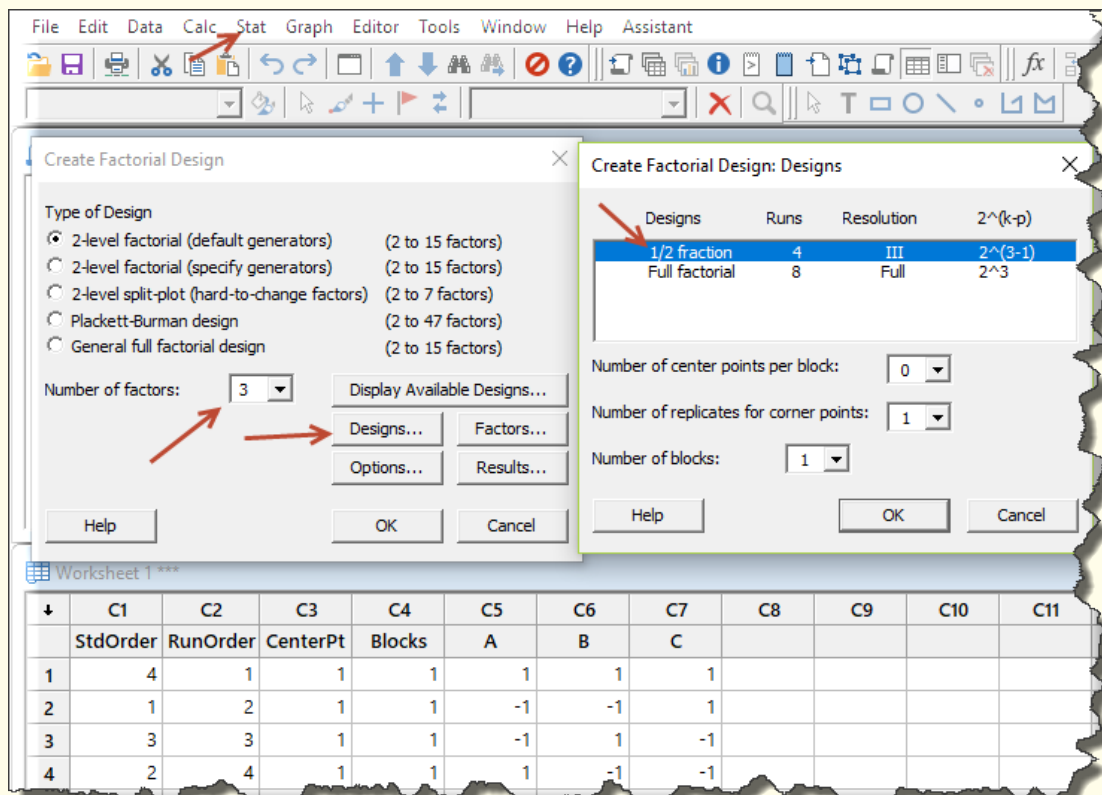
Example 8.1 (1/2 fractional factorial design)

A process has three factors A , B , and C , and the manager does not have resources and time to run all $2^3 = 8$ trials. After applying the 1/2 fractional factorial design, the manager has decided to run $2^{3-1} = 2^2 = 4$ trials where ABC is positive. In other words, the manager will run experiments when the product of the signs in a row is positive. In reference to Table 8.2, runs 2, 3, 5, and 8 qualify for this design. Alternatively, the manager could run trials where the product of the signs in a row is negative. This design corresponds to runs 1, 4, 6, and 7 in Table 8.2.

How-To 8.1 (Factorial design in Minitab 18)

Click on *Stat > DOE > Factorial > Create Factorial Design > Number of factors > Designs > 1/2 fraction > OK*. See the snapshot in Figure 8.1.

Figure 8.1: A 1/2 fractional factorial design of a 2^3 model in Minitab 18. To run all $2^3 = 8$ trials, you would have to click on *Full factorial* in the *Designs* window.



8.2.4 Other considerations of factorial designs

Center points

The factorial designs that we have considered so far assume linearity. In a situation where linearity does not hold, center points can be added to account for non-linearity effects.

Blocking

We employ the blocking option when we have to run experiments under different conditions, such as when we have to run experiments on different work shifts. For example, in a 2^2 factorial design with three replicates, we can run the first set of experiments on the first shift, the second set of experiments on the second shift, and the third set of experiments on the third shift. The three shifts here indicate the number of blocks.

No significant factor

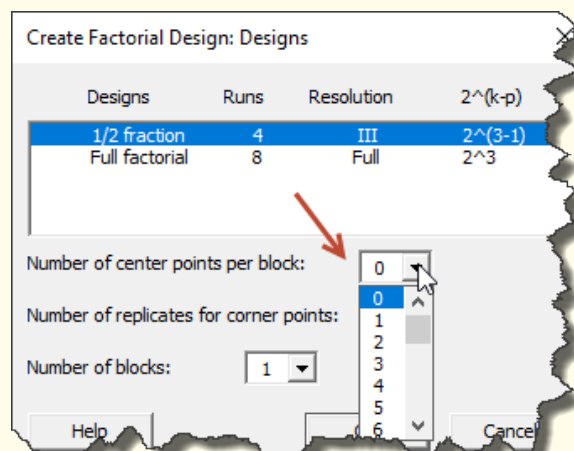
When no effect is statistically significant, we construct a probability plot and then look for points that seem to fall far from the fitted line. Factors corresponding to these points are likely influential in some way. In such instances, we can hypothesize that these outlier points are significant. By pulling the sum of squares of the rest of the factors into the error term, we can apply the F-distribution to test our hypothesis [43].

How-To 8.2 (Center points, replicates, and blocks in Minitab 18)

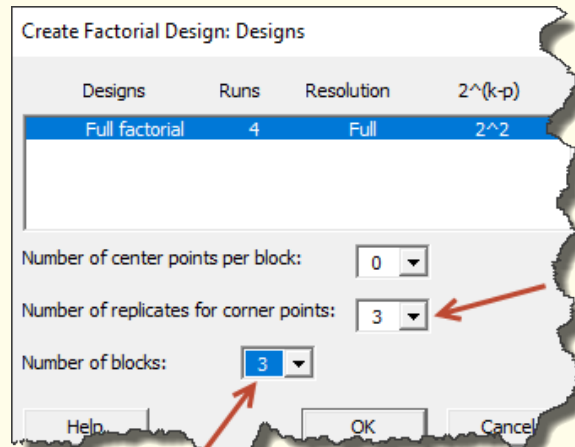
Minitab has an option to add center points as portrayed in Figure 8.2.



Figure 8.2: Center points in Minitab 18 for a $1/2$ fractional design.



See the snapshot in Figure 8.3 for an illustration of the options for replicates and blocks in Minitab 18.

Figure 8.3: Options for replicates and blocks in Minitab 18**Example 8.2 (Factorial design in a HIM department, scenario 1)**

The director of the health informatics department (HIM) at Metropolis Hospital is a member of the revenue cycle steering committee. Last week, this committee was brainstorming about different ways of decreasing the number of unbilled accounts. The HIM manager suggested working with physicians to reduce the cycle time of coding queries ^a. The committee endorsed this idea, and the HIM manager was put in charge of the plan.

To get started, the HIM manager designed experiments, as portrayed in Table 8.3. The objective was to find a query technique that physicians could respond to the fastest.

Table 8.3: The design of experiments for the medical query process

Source of query	Low	High
Coders (A)	Electronic query (-1)	Paper query (1)
CDI^b team (B)	Clinical floor (-1)	Call office (1)

The manager ran experiments per specifications in Table 8.4.

Table 8.4: The order of runs

Run	Output	A	B
1	(1)	-1	-1
2	<i>a</i>	1	-1
3	<i>b</i>	-1	1
4	<i>ab</i>	1	1

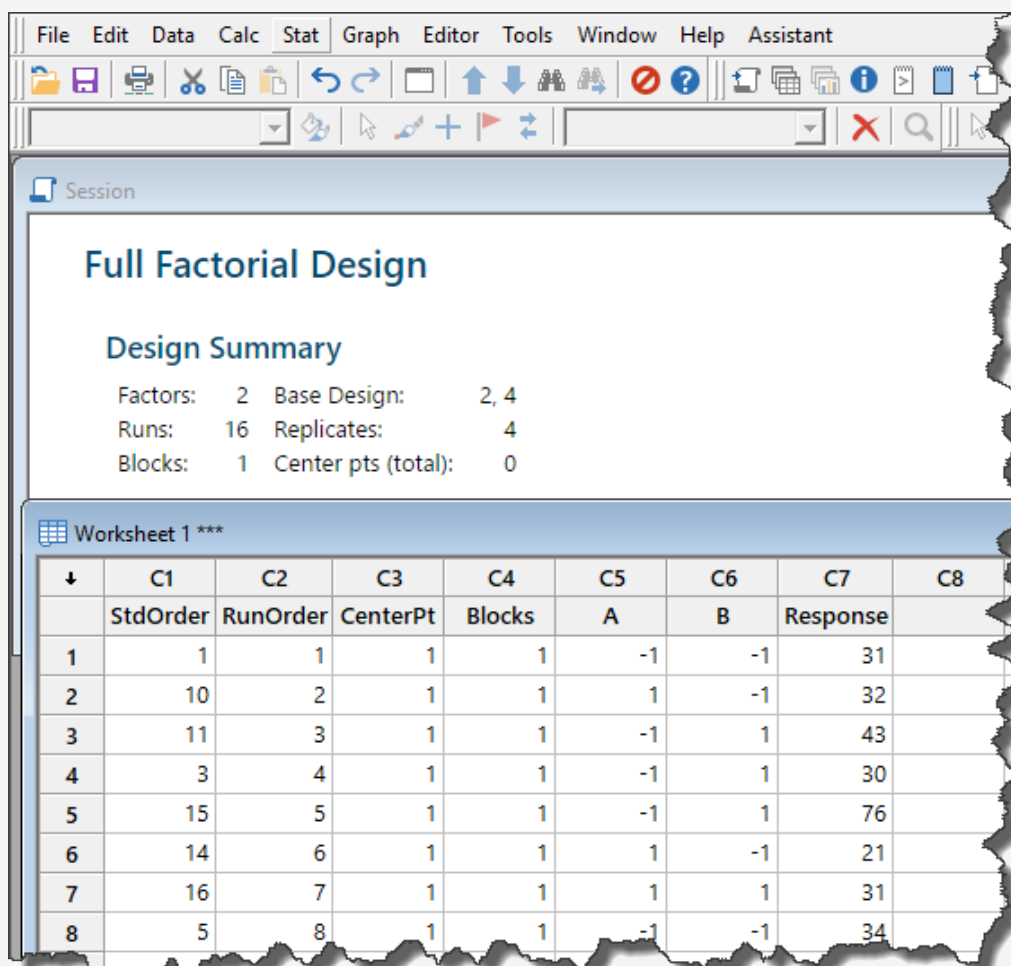
Moreover, the manager replicated the experiments 4 times (R_1, \dots, R_4). The results, which are the averages of the cycle time in hours, are recorded in Table 8.5.

Table 8.5: Experimental results from the medical query process

Run	Output	A	B	R1	R2	R3	R4	Total
1	(1)	-1	-1	31	34	30	33	128
2	a	1	-1	32	21	37	29	119
3	b	-1	1	43	30	76	80	229
4	ab	1	1	31	100	27	99	257

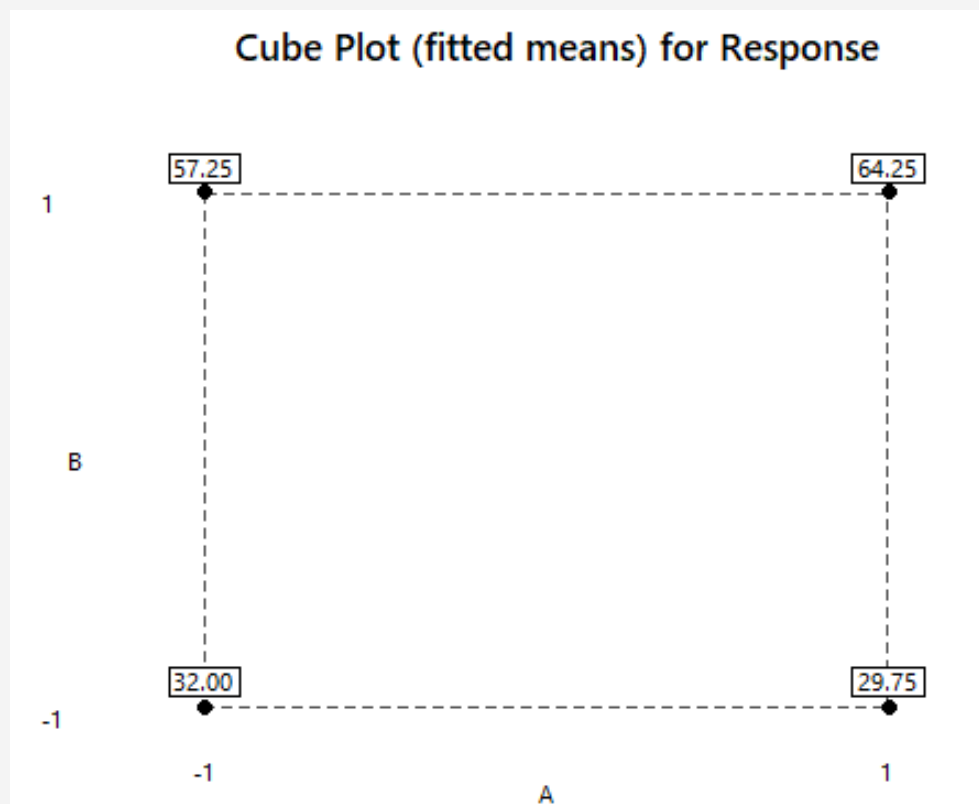
To help the manager analyze these results in Minitab, follow these steps:

1. Open Minitab and design a 2^2 experiment with four replicates. To do this, click on *Stat > DOE > Factorial > Create Factorial Design*. Keep the default settings. After clicking on the Designs tab, change the number of replicates to 4. No center or blocks are considered here.
2. Create a *Response* column and input the data from Table 8.5 as illustrated in the snapshot in Figure 8.4. Make sure you match the signs correctly.

Figure 8.4: Recording Table 8.5 experimental results in Minitab 18

3. To create a cubic plot of your design, click on *Stat > DOE > Factorial > Cube plot*, and your graph should look as depicted in Figure 8.5. The boxed numbers in the corners of this figure represent averages of runs in each row, or simply the total column divided by 4. For example, 32 (in the left lower corner) was obtained by $128/4 = 32$.

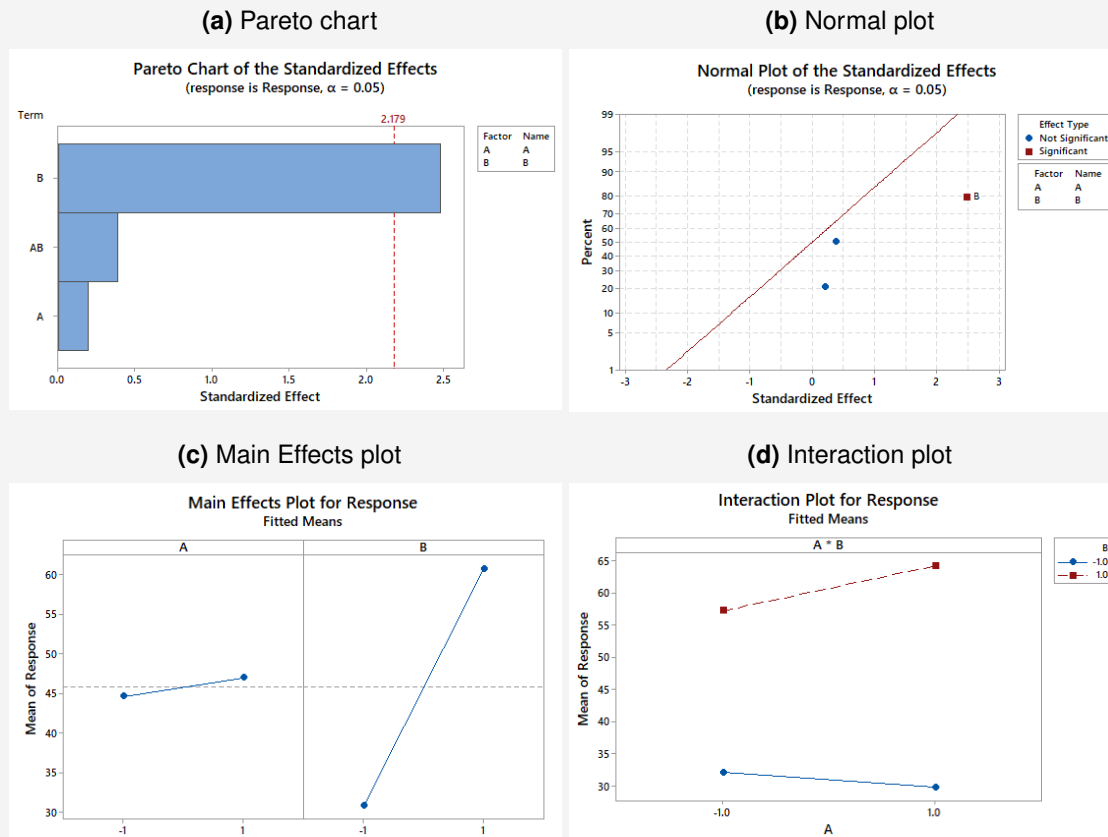
Figure 8.5: A Minitab cubic plot of Table 8.5 experimental results



From the cubic plot in Figure 8.5, we can conclude that when *B* is set to low (-1), the average query response time changes very little when *A* is varied (32 versus 29.75). The same conclusion appears to be true when *B* is set to high (1), and *A* is varied; 57.25 is not that far from 64.25. In other words, the setting of *A* has little effect in this experiment. The only setting that seems to matter is that of *B*. Notice that when *A* is kept constant, varying *B* yields very different results. To test the statistical significance of the setting of *B*, we need to run ANOVA.

4. To view the ANOVA test and other factorial results, click on *Stat > DOE > Factorial >> Analyze Factorial Design*, then select the response column. Click on the *Graphs* tab and ensure that the Pareto chart and the Normal plot are selected. You can view additional factorial plots by clicking on *Stat > DOE > Factorial > Factorial plots > OK*. Your factorial plots may look like the charts displayed in Figure 8.6.

Figure 8.6: Factorial plots from Minitab 18 based on the experimental results in Table 8.5



From Subfigure 8.6a, we can conclude that factor B is statistically significant at the significance level of $\alpha = 0.05$. Our conclusion is based on the fact that the vertical red dotted reference line (2.179) is crossing the B bar [35]. We can reach the same conclusion by also looking at the Normal plot in Subfigure 8.6b, where Minitab identified factor B as being significant. The ANOVA results in Table 8.6 confirm that factor B is indeed statistically significant since the p-value is less than $\alpha = 0.05$.

- To determine the best settings of the new query process, we refer to Subfigure 8.6c. We ignore Subfigure 8.6d since there is no statistically significant interaction between factors A and B (the lines don't cross). From Subfigure 8.6c, it is evident that the average cycle time is minimized when B is set to -1. From Table 8.3, we appreciate that the setting of B = -1 corresponds to having the CDI team query physicians on clinical floors. So, the conclusion is that **moving forward, the CDI team will handle all medical queries on the clinical floors. All other forms of queries are not likely to make much difference.**

Table 8.6: ANOVA test results from Minitab 18 based on the experimental data in Table 8.5

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	3	3678.2	1226.06	2.12	0.151
Linear	2	3592.6	1796.31	3.11	0.082
A	1	22.6	22.56	0.04	0.847
B	1	3570.1	3570.06	6.17	0.029
2-Way Interactions	1	85.6	85.56	0.15	0.707
A*B	1	85.6	85.56	0.15	0.707
Error	12	6938.2	578.19		
Total	15	10616.4			

A Minitab regression model related to this case follows.

$$Y = 45.81 + 1.19A + 14.94B + 2.31A * B \quad (8.21)$$

We can manually recreate this model using Equations 8.1 - 8.5 as follows:

$$\beta_0 = \text{average}(\text{runs}) = 45.8125 \quad (8.22)$$

$$E_1 = \frac{119 + 257 - 229 - 128}{2 * 4} = \frac{19}{2 * 4} = 2.375 \quad (8.23)$$

$$E_2 = \frac{229 + 257 - 119 - 128}{2 * 4} = \frac{239}{2 * 4} = 29.875 \quad (8.24)$$

$$E_{12} = \frac{257 + 128 - 119 - 229}{2 * 4} = \frac{37}{2 * 4} = 4.625 \quad (8.25)$$

Subsequently, we obtain:

$$Y = 45.8125 + \frac{2.375}{2}A + \frac{29.875}{2}B + \frac{4.625}{2}AB + \epsilon \quad (8.26)$$

$$= 45.81 + 1.19A + 14.94B + 2.31A * B + \epsilon \quad (8.27)$$

Holding all other factors constant, the interpretation of this regression model is that increasing B by one unit will decrease Y by 14.94 hours, since when B is set to -1, we get $14.94 * -1 = -14.94$. We should mention that, despite that the $R - sq(adj)$ of this regression model is only 18.31%, this does not negate the fact that factor B is still statistically significant.

^a**Medical query:** a process of asking a medical provider to clarify clinical documentation in health records to promote compliant documentation [49]

^b**CDI:** clinical documentation improvement

Remark: When the interaction of factors is statistically significant, it does not matter whether individual factors in the interaction are statistically significant or not. Only the interaction factor is considered when setting the new process.



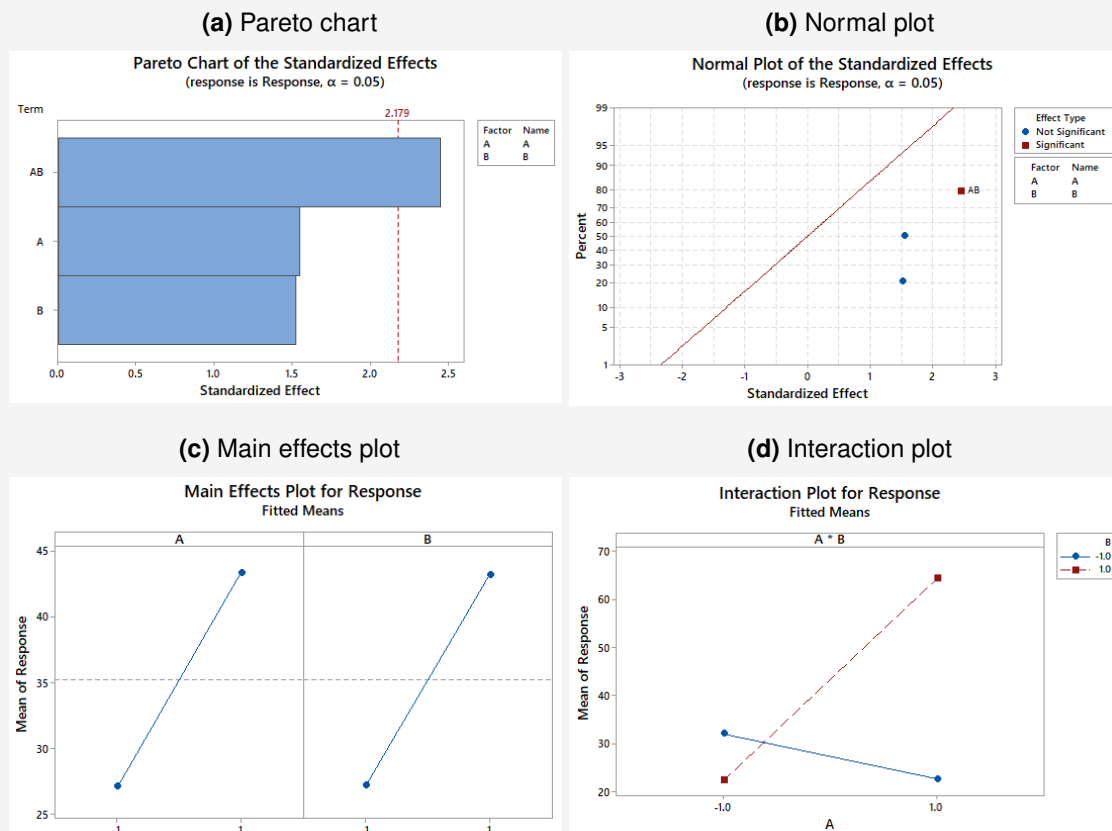
Example 8.3 (Factorial design in a HIM department, scenario 2)

Following from Example 8.2, let's assume that the experimental results were instead as presented in Table 8.7.

Table 8.7: New experimental results of the medical query process

Run	Output	A	B	R1	R2	R3	R4	Total
1	(1)	-1	-1	31	34	30	33	128
2	a	1	-1	32	21	17	20	90
3	b	-1	1	30	12	25	22	89
4	ab	1	1	31	100	27	99	257

Minitab graphical results corresponding to this new scenario are portrayed in Figure 8.7. From these findings, we can conclude that the interaction between factors A and B is statistically significant. As previously remarked about significant interaction terms, in this case, we must set both factors A and B **simultaneously** to minimize Y. From Subfigure 8.7d, the best setting is A set to low (-1), and B set to high (1). That is, **coders will submit an electronic version of the query, and the CDI team will call the physician's office about the same query.**

Figure 8.7: Factorial plots based on the new experimental results in Table 8.7

8.3 Lean tools

The primary goal of *Lean* is to make processes more efficient by eliminating wasteful or non-value added (NVA) activities that increase cycle time.

8.3.1 The cycle time problem

Examples of non-value added activities that increase cycle time include the following **setup problems** [24]:

1. **Learning curves** which occur when we have to perform new tasks or functions that we rarely do.
2. **Interruptions** which happen when we multitask or when we stop what we were doing to track information or wait for approvals. Interruptions not only increase cycle time but also create opportunities for *storage errors*, a recipe for nonconforming services.
3. **Batch processing** such as when we let work accumulate to a certain level before we take action. This practice has the potential for increasing cycle time by delaying the start of the next task in the process.

Other process issues that tend to **increase cycle time** include [24]:

1. **Overprocessing**, which occurs when we try to add more value to service or a product more than the customer requires (e.g., excessive inspection, approvals, and handoffs).
2. **Avoidable transportation and motions**, which include instances of people physically or virtually moving to look for, collect, or deliver materials. Another example is when people perform needless computer keystrokes and tasks.
3. **Overproduction, too much inventory, and excess in work-in-process (WIP)**, which, by Little's law [38], increase cycle time.

8.3.2 Potential solutions

Learning curves

We expect **learning curves** to decrease in the long run. So, if we created a scatter plot of the learning curve over time, we should observe negatively sloped or exponentially decaying values. We can test the significance of this decay statistically [61]. Failure to detect any significant decrease may be a sign of a poor process or lack of required skills to perform the tasks under consideration. Possible solutions include redesigning the process **to reduce complexity**, increasing training, or assigning difficult tasks to **highly-skilled employees**.

To eliminate **interruptions**, we should attempt **offloading** preparatory work by completing interruptive tasks outside of the main process. We should also try to **automate** tasks that computers could do better. Jobs that we cannot offload or automate should be as lean as possible with **no batch processing** unless required.

Example 8.4 (Offloading in medical coding and abstracting process)

There is an opportunity for offloading medical abstracting^a when this task is combined with the main function of medical coding^b. One way to structure such offloading is by implementing at least one of the following routines:

1. Automate the abstracting task such as in the electronic health records (EHR) system
2. Have someone else, besides the coder, do the abstracting job.
3. Have all the information that needs abstracting ready for the coder before coding (e.g., information printed out or in one location in the EHR)

Offloading the abstracting task is likely to decrease the mental setup and coding storage errors due to multi-tasking.

^a**Medical abstracting** is a process in health information management (HIM) for reporting the patient's administrative information such as the discharge disposition and specific quality measures.

^b**Medical coding** is a process of HIM that involves codifying medical diagnoses and procedures into alphanumeric codes using national and international standards.

Overprocessing

To solve the issue of **overprocessing**, we should examine all handoffs, inspections, and rework to ensure that they are necessary. Suitably, we should **empower employees to perform tasks without needing too many approvals**.

Unnecessary transportation

Lean approaches to solving the issues of unnecessary transportation and motions include **rearranging the workspace** to match the flow of work. Additionally, we should redesign computer user interfaces to make them more friendly. We could further streamline the workflow by anticipating the needs of the user and running computer subroutines offline to produce needed information on demand [24]. Another effective solution for avoiding unnecessary movement and motions is **5S+1**, which entails **cleaning our workspace**. The primary goal of **5S +1** is reducing the time that we spend looking for materials and information. The principal components of 5S +1 include[24]:

1. **Sort**: cataloging materials into the categories of the needed to the unneeded.
2. **Straighten**: arranging materials in familiar places.
3. **Shine**: keeping the workstation clean.

4. **Standardize**: maintaining the first 3S's
5. **Sustain**: having a continuous discipline to preserving the workspace procedures.
6. **Safety**: removing all hazardous materials from the workspace.

Overproduction, too much inventory, and excess WIP

Queueing theory can help regulate the rates of production, inventory, and WIP. One caveat is that queueing methods tend to be more technical, unlike standard lean solutions, and may require specialized mathematical training to implement them correctly. But, there are well-known queueing results that we can implement with little efforts, such as:

1. **Pooling**: by cross-training employees, so when demand is high, we have enough resources to address the surge in demand [27].
2. **Triaging**: by sorting jobs into various levels of urgency and processing times. We could minimize the overall cycle time by first processing jobs that require the least amount of time [27].
3. **Optimal staffing** by allocating enough resources to meet the demand [47, 46, 30].

8.4 Data quality tools

One of the major roles of a health informatician is promoting data quality by creating and managing an effective information system. At the minimum, the following design should be enforced to ensure the quality of health care data [61]:

1. **A single point of data collection at the point of entry** to minimize redundancy and increase the accuracy and completeness of the data.
2. **Standards for merging data from multiple sources** for reporting purposes. Examples include definitions of common classes of outcomes, common metrics, standard analytical methods, and patient registries.
3. **Safeguards to ensure security and confidentiality** of patient data such as implementing the provisions of HIPAA²

Some of the necessary steps to improving data quality in any information system include [28]:

1. **Conducting regular audits** to detect nonconforming practices of data capture and use.

²**HIPAA**: Health Insurance Portability and Accountability Act

2. **Establishing an information governance structure** in the organization and getting everyone committed to quality data.
3. **Improving data capture controls** since when poor data are captured, only poor data could be made available for decision making (e.g., GIGO³).
4. Keeping up with modern **data management principles**
5. **Involving executives** and proving to them the value of data quality

8.5 Prevention tools

There are no better quality improvement techniques than the ones that prevent quality problems from happening in the first place.

8.5.1 Poka-Yoke measures

Poka-yoke⁴ measures are mistake-proofing safeguards that we put in place to mitigate the occurrence of adverse events. In general, it is difficult to come up with such measures since one cannot anticipate all future sources of quality problems. The collective ingenuity of process stakeholders, experts, and the public is a must. The website of the Agency for Healthcare Research and Quality (AHRQ) [3] <https://archive.ahrq.gov/professionals/quality-patient-safety/patient-safety-resources/resources/mistakeproof/index.html> has several examples of error-proofing measures in health care. One of these examples is a Computerized Physician Order Entry (CPOE) system to mitigate medication errors [34].

8.5.2 Swiss-cheese model

Many quality problems, especially complex ones, are not amenable to poka-yoke prevention measures. For such issues, we should at least apply the **Swiss cheese** model by creating several layers of checkpoints to attempt to catch and fix any anomalies [55]. We must align the Swiss cheese layers correctly so that an anomaly that we missed at the previous inspection, we find it at the next checkpoint. Of course, despite our best efforts, some quality issues will slip through, which justifies the need for continuous quality improvement.

8.6 Control charts

Control charts are tools that we typically use to monitor a process. But, these tools are also critical in improving the process. A typical control chart has a centerline (CL),

³**GIGO**: garbage-in, garbage-out

⁴**Poka-yoke**: a Japanese word that is widely used in Lean circles to indicate mistake-proofing or error-proofing

an upper control limit (UCL), and a lower control limit (LCL). We monitor a process to ensure that it stays within the control limits. An easy way to interpret control charts is to look for points that fall outside of the limits since these points indicate samples with possible assignable causes for out-of-control behaviors. As we discuss more in Module 9, **by removing assignable causes, variability decreases, which in turn improves the process** [43]. Indeed, a process with no special cause variation needs no improvement.

8.7 Check your understanding

- At which phase of DMAIC should we implement the poka-yoke measures?
- During your DOE analysis, you found the effect of factor X to be 10. What is the related coefficient of regression?
 - 5
 - 10
 - 15
 - 20
- An ANOVA test from designed experiments shows the following results:

Table 8.8

Source	SS	DF	MS	F	P
A	15	1	15	1.5	0.5
B	80	2	40	4	0.01
A*B	30	2	15	1.5	0.1
Error	180	18	10	1	
Total	305	23			

- What can you conclude about DOE results in Table 8.8?
 - Based on this table, how many replicates were there? *Hint: think in terms of the degrees of freedom and the levels of a factor*
- Consider the experimental results in Table 8.9.

Table 8.9

Run	Output	A	B	R1	R2	R3	R4	R5
1	(1)	-	-	100	83	77	74	71
2	a	+	-	70	44	53	48	58
3	b	-	+	31	36	28	36	26
4	ab	+	+	98	73	45	50	58

Using this data,

- Run DOE analysis in Minitab
- Recreate your regression model in Excel

- (c) What settings are you going to adopt for the new process?
 - (d) What concerns do the residuals plots reveal?
 - (e) Attach both your Excel and Minitab files along with the interpretation of the:
 - i. ANOVA results
 - ii. Pareto, Normal, Main Effects, and Interaction plots
 - iii. Regression model
5. From your DOE study with three main factors (A, B, and C), you found A, B, A*B, and A*B*C to be statistically significant. The new process should be redesigned based on the settings of:
- (a) factor A only
 - (b) factor B only
 - (c) factor A*B only
 - (d) factor A*B*C only
 - (e) None of the above
6. From your DOE analysis, you found no statistically significant factor. What typical graphical tool can you utilize to spot factors that may still be significant after adjusting the error term?
- (a) Scatter plot
 - (b) Normal plot
 - (c) Main effect plot
 - (d) Interaction plot
7. Your original factorial design was 2^4 . If you decided to run a $1/2$ fractional design instead, how many experiments would you run?
- (a) 4
 - (b) 8
 - (c) 16
 - (d) 32
8. A more general form of a fractional factorial design with two levels is:
- (a) 2
 - (b) 2^k
 - (c) 2^{k-1}
 - (d) 2^{k-p}
9. Given $SS_A = [a + ab - b - (1)]^2/4n$, what is the contrast?

10. What is the main purpose of 5S +1?
11. Why is the **Swiss cheese** model applicable for quality improvement in health care informatics?
12. When do we apply **DOE** tools, and why?
13. **Queueing theory** is applied in quality improvement to help address which type of process issue?
 - (a) Overproduction
 - (b) Overprocessing
 - (c) Learning curves
 - (d) Interruptions
14. **Eliminating unnecessary approvals and signatures** will likely help us reduce NVA time due to:
 - (a) Overproduction
 - (b) Overprocessing
 - (c) Learning curves
 - (d) Interruptions
15. Assigning difficult tasks to **highly-skilled employees** will likely help us mitigate NVA time due to
 - (a) Overproduction
 - (b) Overprocessing
 - (c) Learning curves
 - (d) Interruptions
16. **Offloading** will probably help us reduce NVA time due to:
 - (a) Overproduction
 - (b) Overprocessing
 - (c) Learning curves
 - (d) Interruptions
17. Always keeping the **workstation clean** relates to which element of 5S +1?
 - (a) Sort
 - (b) Straighten
 - (c) Shine
 - (d) Standardize

18. Which tool is likely to help us best manage a queue from **a sudden spike in demand**?
- (a) Pooling
 - (b) Triaging
 - (c) Offloading
 - (d) 5S+1
19. **Triaging and prioritizing** shorter processing time jobs is likely to help us minimize the:
- (a) Lead time
 - (b) Completion rate
 - (c) Arrival rate
 - (d) NVA time
20. From your DOE analysis, you found the interaction between factors A and B to be statistically significant, and your new process will involve the simultaneous setting of $A = -1$ and $B = +1$. If your regression equation was as follows:

$$Y = 10 + 7A + 3B - 17AB$$

what can you conclude about the coefficients of the model?

Tools for monitoring and controlling the process

In this module, we review techniques for monitoring and controlling a process. Some of the tools that we consider include control and run charts. We also discuss the implementation strategies of control charts and the design of an out-of-control action plan (OCAP) to regulate the future process. For software, we utilize Minitab and Excel.

Key concepts and tools: Shewhart model; Control charts; Run charts; Out-of-control behaviors; Assignable causes; Sensitizing rules; Special cause variation; Common cause variation; Sample size and frequency; Average run length (ARL); Phases of control charts; Variable charts; Attribute charts; OCAP; Dashboards

Major objectives

After studying this module, you will be able to:

- 1. Define key concepts and tools for monitoring and controlling a process*
- 2. Justify the need for monitoring variability in the process*
- 3. Explain the Shewhart control chart model*
- 4. Implement \bar{X} and p control charts*
- 5. Implement run charts*
- 6. Differentiate the phases of chart application*
- 7. Identify assignable causes in the process*
- 8. Implement dashboards*
- 9. Create an OCAP*
- 10. Advocate for quality improvement using control charts*

9.1 Introduction

After we have improved and stabilized our process, we proceed to the *Control* phase of DMAIC, where we implement monitoring measures to detect any new out-of-control behaviors in the process. Some of the tools that we employ include control and run charts. During this phase, we also create an OCAP and a transfer plan of the process.

9.2 Control charts

Control charts are essential tools of SPC¹ that we utilize to monitor and improve quality. There are two main types of control charts (**Shewhart** and **time-weighted**) that we implement in two phases. In **phase I**, we implement Shewhart charts to detect **big shifts** in the process. Phase I is essentially the period when we attempt to stabilize the process (e.g., during the *Improve* phase of DMAIC). In **phase II**, we implement time-weighted charts to monitor **small shifts** in the process (e.g., during the *Control* phase of DMAIC). Although, we may still be able to employ Shewhart charts in phase II using rules known as **sensitizing**, as we demonstrate later. The three common types of time-weighted control charts that do not discuss further here are Cumulative Sum (**CUSUM**), Exponentially Weighted Moving Average (**EWMA**), and Moving Average (**MA**) [43]. In this module, we only consider the application of Shewhart control charts.

9.2.1 The Shewhart model

Walter A. Shewhart is credited for introducing the concept of control charts in the 1920s while working at Bell Labs [43]. As summarized in Box 9.1, the Shewhart's model of a control chart has three main components: an upper control limit (*UCL*), a centerline (*CL*), and a lower control limit (*LCL*).

Box 9.1 Shewhart's model of control charts

The construct of a Shewhart control chart is as follows:

$$UCL = \mu_w + L\sigma_w \quad (9.1)$$

$$CL = \mu_w \quad (9.2)$$

$$LCL = \mu_w - L\sigma_w \quad (9.3)$$

where w is the statistic of interest. Here, μ_w and σ_w symbolize the mean and standard deviation of w , respectively. Letter L denotes the distance measured in standard deviation units from CL to LCL or UCL [43].

Control charts built following Box 9.1 are referred to as **Shewhart control charts**.

¹SPC: Statistical Process Control

9.2.2 How do Shewhart control charts work?

An analogy that is often used to describe the mechanics of a Shewhart control chart is that of *hypothesis testing* (review Module 7). For demonstration purposes, let's consider the following hypothesis of a normally distributed process:

$$H_0 : \mu = \mu_0 \quad (9.4)$$

$$H_1 : \mu \neq \mu_0 \quad (9.5)$$

Here, μ is the process mean and μ_0 is the hypothesized mean that we estimated from the process samples. The standard deviation σ is given, otherwise, we are able to obtain its unbiased estimator from the process samples. If \bar{x}_i is the mean of sample i of this process, the following test statistic L applies to the hypothesis under consideration:

$$L = \frac{\bar{x}_i - \mu_0}{\sigma/\sqrt{n}} \quad (9.6)$$

where n is the sample size. We obtain the confidence interval of the process mean μ in terms of \bar{x}_i as follows:

$$\mu_0 - L \frac{\sigma}{\sqrt{n}} \leq \bar{x}_i \leq \mu_0 + L \frac{\sigma}{\sqrt{n}} \quad (9.7)$$

We can think of the lower limit of the confidence interval in Equation 9.7 as LCL , the upper limit as UCL , and μ_0 as CL . If \bar{x}_i falls in this interval, then $\bar{x}_i \approx \mu_0$ and we conclude that sample i is in control. If $\bar{x}_i \neq \mu_0$, \bar{x}_i will fall outside of the confidence interval, which will lead us to reject H_0 . Accordingly, we will conclude that the process is **out-of-control** or has **special cause variation** at that particular sample number. Failure to reject H_0 at any sample number, is equivalent to saying that the process is in **statistical control**, or simply that the process only exhibits **common cause** variation. By common cause variation, we imply that the process **behaves randomly** within the control limits.

9.2.3 Implementing Shewhart control charts

Shewhart control charts are effective at improving quality if they are implemented correctly. For example, before implementing control charts, we must decide on how big the sample sizes must be and how frequently sampling should be done. Besides, we must understand how to choose distance L and know how to detect out-of-control behaviors.

Sampling

Adequate **sample sizes** and the **frequency** of sampling are required for a successful application of control charts. In general, we need at least 25 samples to construct a control chart. The larger the sample sizes, the easier it is to detect small shifts in the process. If obtaining large samples is not feasible, we try to take small samples but do it more frequently. In this effort, we must ensure **rational subgroups** to maximize the between samples differences while minimizing the within-sample differences. Doing so

will promote the independence of the samples, an assumption that we often make when creating Shewhart control charts. An appropriate partition of the **time-ordered** samples is critical in minimizing the dependence of the subgroups [43]. In general, sampling only from units produced since the last subgroup, or from units produced under similar conditions, will help ensure rational subgroups [43].

Choosing L

As indicated earlier, letter L in a Shewhart control chart denotes the distance, in standard deviation units, from the centerline to the control limit. It is customary to choose $L = 3$, but depending on the application, this setting may be different. For example, in the type of control charts known as *risk-adjusted*, the limits are typically set to $L < 3$ [61]. Why should we care about where we set the control limits? We should care because different settings of L may lead to different conclusions about the following hypothesis:

$$H_0 : \text{process in control} \quad (9.8)$$

$$H_1 : \text{process not in control} \quad (9.9)$$

As it can be observed in the confidence interval in Equation 9.7, if L is too big, meaning too far from CL , \bar{x}_i is more likely to fall within limits. This outcome increases the probability β for **Type II error**, meaning that we are likely to fail to reject H_0 when it is false. In contrast, if L is too small, \bar{x}_i is more likely to fall outside of the limits, which increases the probability α of **Type I error** where we reject H_0 when it is true. We can think of type I error as a **false alarm** in a stable process [43, 61]. Both type I and II errors affect how we measure the performance of control charts. For example, using the probability of type I error α , we can construct the control chart performance measure referred to as **average run length (ARL)**, as follows:

$$ARL_\alpha = \frac{1}{\alpha} \quad (9.10)$$

The ARL_α measure also known as *in-control ARL* or ARL_0 , allows us to determine the expected number of samples until a false alarm occurs in a stable process. In a normally distributed process with $L = 3$, the probability of a point falling either outside of LCL or UCL is given by $1 - 0.9973 = 0.0027$. This implies that $ARL_\alpha = 1/0.0027 \approx 370$. Hence, we can expect a false alarm at every 370th sample, if the process has not shifted. If the process mean has shifted, we gauge the process performance using the ARL_β measure that we obtain like this:

$$ARL_\beta = \frac{1}{1 - \beta} \quad (9.11)$$

As previously noted, β is the probability of type II error, and, as discussed in Module 7, $1 - \beta$ is the power test that expresses the probability of correctly rejecting H_0 . The ARL_β measure in Equation 9.11 is also referred to as *out-of-control ARL* or ARL_1 . This statistic expresses the expected number of samples to detect a mean shift of $k\sigma$ in the process. In practice, we want to be able to detect a shift in the process as soon as it happens. We can accomplish this by increasing the sample size n since that decreases β [43].

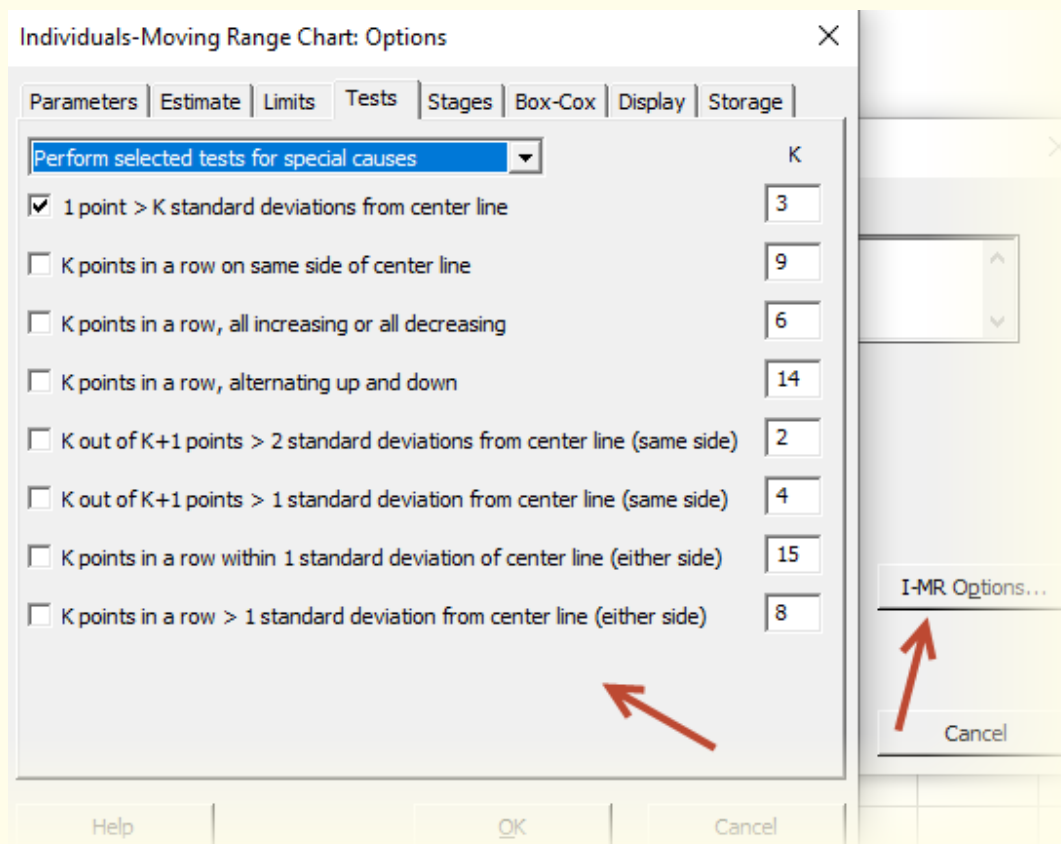
Sensitizing rules

We have previously noted that when all sample statistics of interest fall within the control limits, the process is in statistical control. We want to underscore that this conclusion is true only when points fall within the control limits randomly. Otherwise, if a particular pattern exists, out-of-control behaviors are likely due to **small shifts** in the process. One way to detect such shifts is to employ specialized rubrics referred to as **sensitizing rules**, as exemplified in How-To 9.1. It is important to emphasize that these rules are not universal. Another way to detect small shifts in the process is by utilizing **time-weighted** control charts.

How-To 9.1 (Sensitizing rules in Minitab 18)

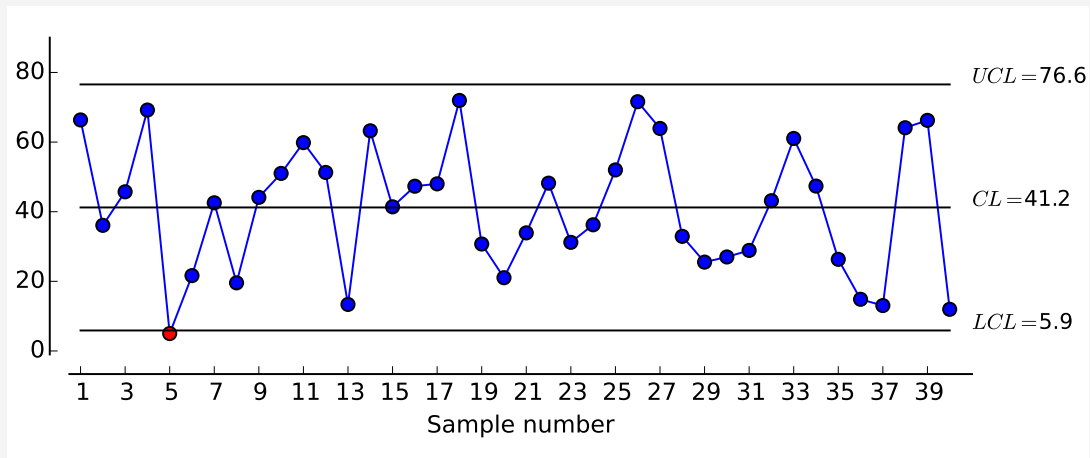
For each control chart, Minitab 18 displays applicable sensitizing rules. To access these rules, click on *Chart options* > *Tests*. See Figure 9.1 for an example of sensitizing rules of an *ImR* control chart.

Figure 9.1: Sensitizing rules of an *ImR* control chart in Minitab 18

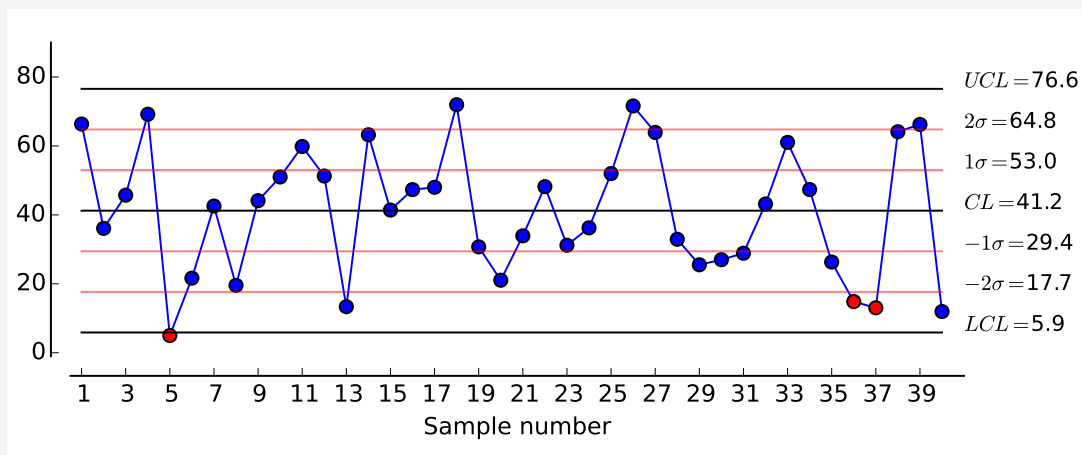


Example 9.1 (Special cause variation)

Figure 9.2 shows an example of a control chart with out-of-control behaviors since at sample #5, a point fell outside of *LCL*.

Figure 9.2: A control chart with one type of special cause variation

After applying the sensitizing rules, Figure 9.3 further indicates that this process has special cause variation at samples 35 and 36, given that 2 out of 3 consecutive points are between the second and third standard deviation limits on the same side.

Figure 9.3: A control chart with two types of special cause variation

9.2.4 How to improve a process using Shewhart control charts?

A process with out-of-control behaviors is **unstable**. To stabilize the process, we must attempt to assign a reason to the special cause and try to rectify it. When the cause is **assignable** and rectifiable, we remove the out-of-control sample from the process and recalculate the centerline and control limits. **By removing assignable causes, variability decreases, which in turn improves the process** [43]. A process with no special cause variation needs no improvement. But, one could still decide to **reduce common cause variation**, which amounts to creating a new process [43].

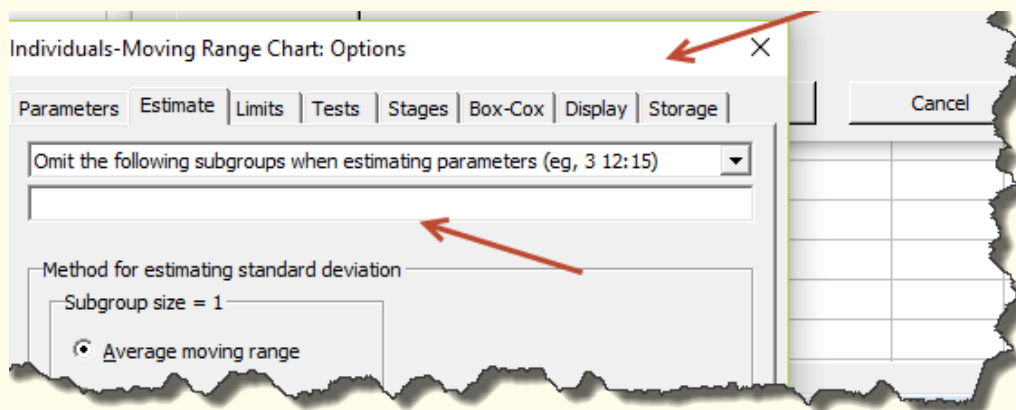


Remark: Control charts only tell us about the samples that are out-of-control but say nothing about the actual reasons for out-of-control behaviors. Assigning causes and fixing particular sources of the special cause variation requires good knowledge of the process, and a well-tuned understanding of how to apply various quality improvement tools, including Lean methods and designed experiments.

How-To 9.2 (Omitting samples in Minitab 18)

To remove an assignable cause, click on *Chart Options* tab of a control chart > *Estimate* > *Enter the sample number(s)* to be omitted. See the snapshot in Figure 9.4.

Figure 9.4: Omitting samples with assignable causes in Minitab 18



Note: After omitting a sample number, Minitab 18 calculates a new centerline and also adjusts control limits, but, depending on setting in Minitab, the omitted sample may still be visible on the chart.

Example 9.2 (Assignable special cause variation)

A medical coding manager at Metropolis Hospital monitors the daily processing time of inpatient records. The control chart indicates that the sample from last Monday was out-of-control. After brainstorming with coders and the IT staff, the manager realized that the reason for out-of-control behaviors was encoder updates that overwrote some coding data on that day. Since the manager knows the reason for the out-of-control behaviors (that is, the manager assigned the cause), the sample from last Monday will be omitted, and the centerline and control limits recalculated. By removing the out-of-control sample, variability will decrease in the process control chart.

Example 9.3 (Common cause variation)

The office manager at Metropolis Radiology Center uses control charts to monitor the average coding time for outpatient records. No special variation causes have been noticed, but still, the manager would like to reduce common cause variation

by creating a new process based on computer-assisted coding (CAC) technology. The manager believes that the new process will minimize variability by improving the consistency of coding. The manager will need new control charts to monitor the new process.

9.2.5 The types of Shewhart's control charts

The two common categories of Shewhart's control charts are **variable** and **attribute**. We use **variable** charts when a process generates continuous independent data. Typical variable charts include ImR , $XbarR$, and $XbarS$. We employ ImR charts when the sample size $n = 1$. We apply $XbarR$ charts when $1 < n \leq 10$. When $n > 10$, we use $XbarS$ charts. The normal distribution is typically assumed in all variable charts. We apply **attribute** charts when a process generates independent discrete data. Typical attribute charts include p , np , c , u , g , and h . The p chart relates to the Bernoulli random variable. The np chart relates to the binomial random variable. The c and u assume the Poisson random variable. The g and h charts presume the geometric random variable.

In this module, we only demonstrate how to build ImR and p charts. These charts are likely to be sufficient when monitoring basic processes in health care informatics. For the reader interested in learning how to construct other types of Shewhart charts, as well as time-weighted charts, see a separate book that we have on the application methods of control charts in health care.

9.2.6 ImR charts

We use ImR control charts, also symbolized as $I-MR$, to monitor a process when the sample size $n = 1$. The part of I monitors individual observations, whereas the part of MR , which stands for *moving range*, monitors variability in the process. We denote observation x in sample i as x_i and calculate the average \bar{x} of all observations m like this:

$$\bar{x} = \frac{1}{m} \sum_{i=1}^m x_i \quad (9.12)$$

We determine MR between two successive observations. As such, MR between x_i and x_{i-1} is given by:

$$MR_i = |x_i - x_{i-1}| \quad (9.13)$$

for $i : 2, \dots, m$, where the $|\cdot|$ symbol signifies the absolute value. We calculate the average of moving ranges, \overline{MR} as follows:

$$\overline{MR} = \frac{1}{m-1} \sum_{i=2}^m MR_i \quad (9.14)$$

We estimate the variability in the process this way:

$$\sigma \approx \hat{\sigma} = \frac{\bar{MR}}{d_2} \quad (9.15)$$

where, from Appendix Table 1, $d_2 = 1.128$ when $n = 2$, since we determine MR using two observations. Box 9.2 summarizes the formulas for ImR control charts.

Box 9.2 Formulas for ImR control charts when $L = 3$

The formulas for ImR charts are as follows:

	I chart	MR chart
UCL	$\bar{\bar{x}} + 3 \frac{\bar{MR}}{d_2}$	$D_4 \bar{MR}$
CL	$\bar{\bar{x}}$	\bar{MR}
LCL	$\bar{\bar{x}} - 3 \frac{\bar{MR}}{d_2}$	$D_3 \bar{MR}$

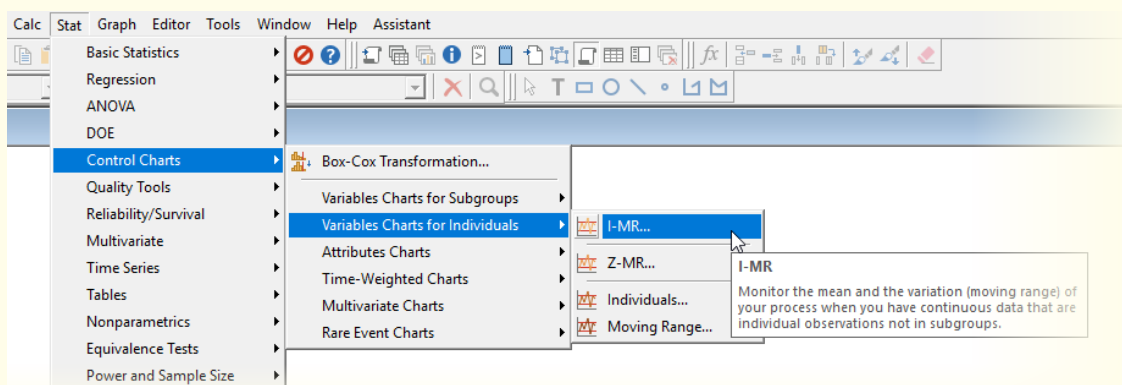
Here, $\bar{\bar{x}}$ is the overall mean of observations. We obtain parameters d_2 , D_3 , and D_4 in Appendix Table 1 when $n = 2$. Since we can't have negative observations, when $LCL < 0$, we set $LCL = 0$.

How-To 9.3 (I - MR charts in Minitab 18)

To construct ImR charts in Minitab 18, click on *Stat > Control Charts > Variable Control Charts for Individuals > I-MR > Select your data > OK*. See the snapshot in Figure 9.5.



Figure 9.5: Options for I - MR control charts in Minitab 18



**How-To 9.4** (*I-MR* charts in Excel 2013)

Excel does not have a built-in option to create *I – MR* charts, but we can manually program the formulas from Box 9.2 in a spreadsheet, as we demonstrate in Example 9.4.

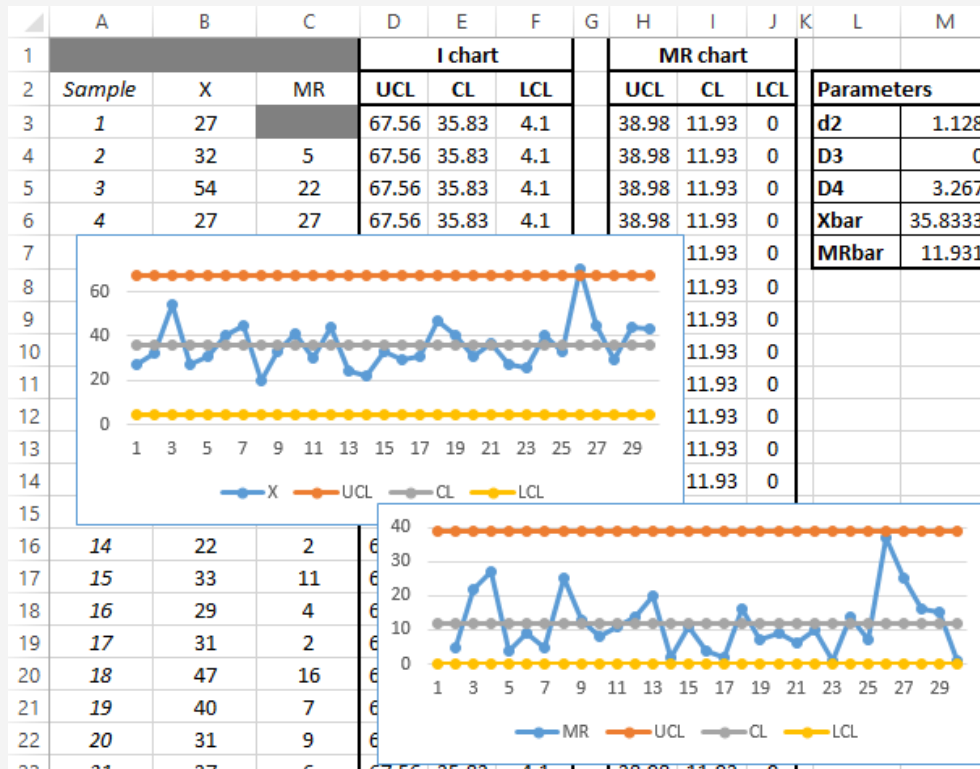
Example 9.4 (*I-MR* charts)

The lab manager at Metropolis Hospital is interested in monitoring the turn around time (TAT) of “stat” potassium orders for each patient. A shorter TAT is preferred since potassium levels are required before patients could be taken off the heart/lung bypass machine in the cardiac unit [50]. To set up preliminary *ImR* control charts, the manager collected 30 random samples, as portrayed in Table 9.1.

Table 9.1: TAT, in minutes (*X*), of “stat” potassium orders at Metropolis Hospital

<i>Sample#</i>	<i>X</i>	<i>Sample#</i>	<i>X</i>
1	27	16	29
2	32	17	31
3	54	18	47
4	27	19	40
5	31	20	31
6	40	21	37
7	45	22	27
8	20	23	26
9	33	24	40
10	41	25	33
11	30	26	70
12	44	27	45
13	24	28	29
14	22	29	44
15	33	30	43

To help the manager construct the appropriate control chart, we start by consulting Appendix Table 1 for $n = 2$ and obtain $d_2 = 1.128$, $D_3 = 0$, and $D_4 = 3.267$. Next, we set up our spreadsheet, as portrayed in Figure 9.6.

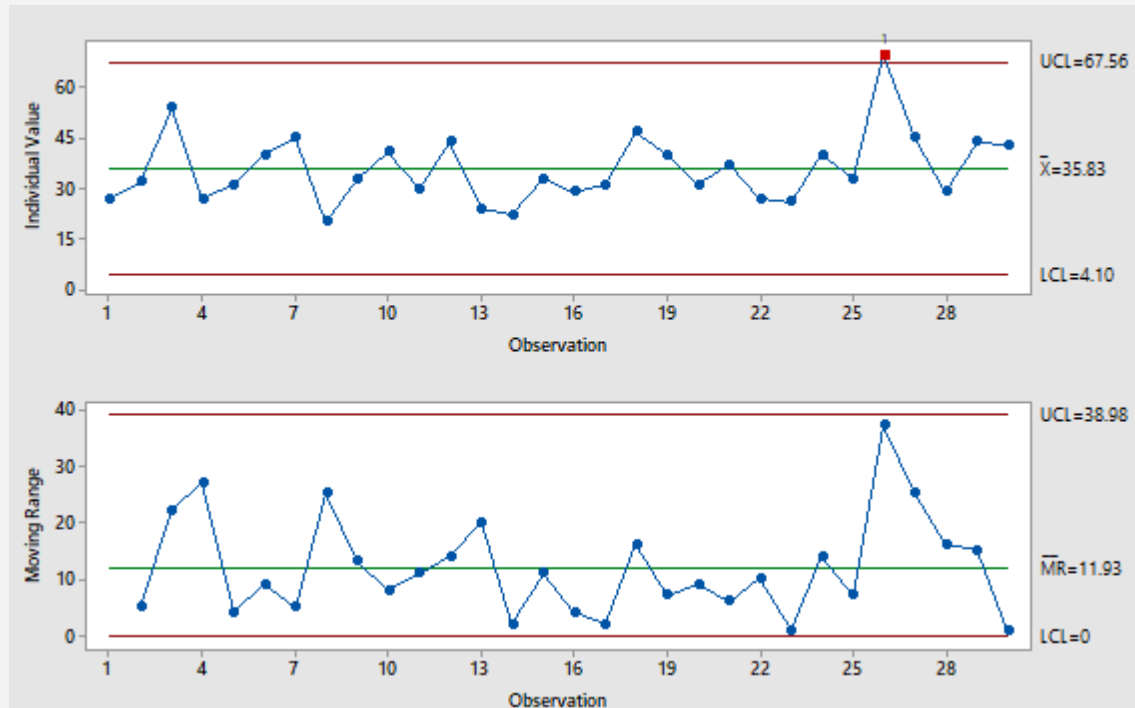
Figure 9.6: A setup of I - MR control charts in Excel based on the data in Table 9.1

To create the I - MR charts, we follow these steps:

- Step 1:** From the setup in Figure 9.6, we have column B containing individual observations that we monitor in the I chart. We also use this column to determine the moving ranges (MR).
- Step 2:** We obtain the MR values per Equations 9.13 and 9.14. For example, we calculated the MR value in cell C4 using $=ABS(B3 - B4)$. We dragged down this formula to populate the rest of the values. Furthermore, we obtained \overline{MR} value in cell M7 using $=AVERAGE(C4 : C32)$.
- Step 3:** In this step, we calculate the \overline{X} value. We obtained the value in cell M6 using $=AVERAGE(B3 : B32)$.
- Step 4:** To construct the control limits, we apply the formulas in Box 9.2. For example, we obtained the UCL value in cell D3 by $=ROUND(\$M\$6 + 3 * (\$M\$7/\$M\$3), 2)$ and obtained the LCL value in cell J3 by $=\$M\$4 * \$M\7 . The CL value in cell E3 follows from $=ROUND(\$M\$6, 2)$. We use the $ROUND(x, 2)$ function to round x to two decimal places. The $\$$ sign in Excel formulas allows us to drag down the formula and easily populate the rest of the values while keeping cells with the $\$$ sign constant.
- Step 5:** In this final step, we insert line charts of columns D-F to create the I chart and line charts of columns H-J to create the MR chart.

Figure 9.7 portrays *I-MR* charts that we created by following the instructions in How-To 9.3.

Figure 9.7: *I-MR* control charts created in Minitab using the data from Table 9.1



From any of these sets of *I-MR* charts, we appreciate that the *I* chart exhibits out-of-control behavior at sample 26. The *MR* chart does not show any special cause variation, but it is also clear that sample 26 is too close to *UCL*. To improve this process, the manager needs to investigate sample 26 and remove any sources of variability. After assigning the causes and fixing the problems, the manager can follow the instructions in How-To 9.2 to omit sample 36, which will result in a decrease in the variability of the process. For example, after trying this, you will notice that in the *I-chart*, *UCL* will decrease from 67.56 to 62.77 and *LCL* will move up from 4.1 to 6.54. In the *MR-chart*, *UCL* will decrease from 38.98 to 34.54. In both cases, the mean values will also decrease. As the new control charts will be stable, after omitting sample 26, the manager will be able to use the revised charts to monitor the future process.

9.2.7 A *p* chart

We use a *p* chart to monitor the fraction of defective units in the process. A *p* chart is based on the Bernoulli random variable given by:

$$p(x) = \begin{cases} p & x = 1 \\ 1 - p & x = 0 \end{cases} \quad (9.16)$$

where p is the probability of an event under study. If a randomly selected unit is defective, $x = 1$, otherwise $x = 0$. The mean μ and the standard deviation σ of the **Bernoulli random variable** are given by:

$$\mu = p \quad (9.17)$$

$$\sigma = \sqrt{p(1-p)} \quad (9.18)$$

Since we generally don't know p , we estimate it from the process samples via the method of the maximum likelihood estimator (MLE). To demonstrate, let's suppose that we took m number of samples with each sample i having a constant size n . If D_i is the number of defective units in sample i , the MLE estimator of p is given by \bar{p} as follows:

$$\bar{p} = \frac{\sum_{i=1}^m D_i}{mn} \quad (9.19)$$

When the sample size is variable, we obtain \bar{p} like this:

$$\bar{p} = \frac{\sum_{i=1}^m D_i}{\sum_{i=1}^m n_i} \quad (9.20)$$

Following from the central limit theorem, we obtain the standard deviation of the sample mean like this:

$$\sigma_{\bar{p}} = \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \quad (9.21)$$

Box 9.3 summarizes the formulas for creating a p chart.

Box 9.3 The formulas for a p chart when $L = 3$

When the **standard** of p is given, we apply these formulas:

$$\begin{array}{l|l} UCL & p + 3\sqrt{\frac{p(1-p)}{n}} \\ \hline CL & p \\ \hline LCL & p - 3\sqrt{\frac{p(1-p)}{n}} \end{array}$$

When **no standard** is given, we apply the following formulas:

$$\begin{array}{l|l} UCL & \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \\ \hline CL & \bar{p} \\ \hline LCL & \bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \end{array}$$

In each sample i , we monitor the fraction of defective units $\hat{p}_i = D_i/n$, when the sample size is constant. When the sample size varies, we monitor $\hat{p}_i = D_i/n_i$ in each sample. The \bar{p} statistic is obtained per Equation 9.19 or 9.20.

Remark ($LCL > 0$): It is common practice to set the LCL value x to $\max(0, x)$ since negative defective units don't make sense. This approach generally works well, but it has a downside since all negative values are treated the same. In other words, we may never know when the process has a very small fraction of defective units since all we may observe is $LCL = 0$. To remedy this problem, we choose the sample size n so that LCL is always positive. To do that, we proceed as follows:



$$\bar{p} - L\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 0 \quad (9.22)$$

$$-L\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = -\bar{p} \quad (9.23)$$

$$\left(-L\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}\right)^2 = (-\bar{p})^2 \quad (9.24)$$

$$\frac{L^2\bar{p}(1-\bar{p})}{n} = \bar{p}^2 \quad (9.25)$$

$$\frac{L^2\bar{p}(1-\bar{p})}{\bar{p}^2} = n \quad (9.26)$$

$$\frac{L^2(1-\bar{p})}{\bar{p}} = n \quad (9.27)$$

So, to ensure that $LCL > 0$, we must choose n such that:

$$n \geq \left\lceil \frac{L^2(1-\bar{p})}{\bar{p}} \right\rceil \quad (9.28)$$

where $\lceil \cdot \rceil$ is a ceiling function.

Alternatively, we could monitor standardized values given by:

$$Z_i = \frac{\hat{p}_i - p}{\sqrt{\frac{p(1-p)}{n_i}}} \quad i = 1, 2, \dots, m \quad (9.29)$$

where $UCL = L$, $CL = 0$, and $LCL = -L$. We recall that \hat{p}_i is the fraction of defective units in sample i and n_i is the corresponding sample size.

How-To 9.5 (A p chart in Excel 2013)

Excel does not have a built-in option to create a p chart, but we can manually program the formulas from Box 9.3 in a spreadsheet, as we demonstrate in Example 9.5.

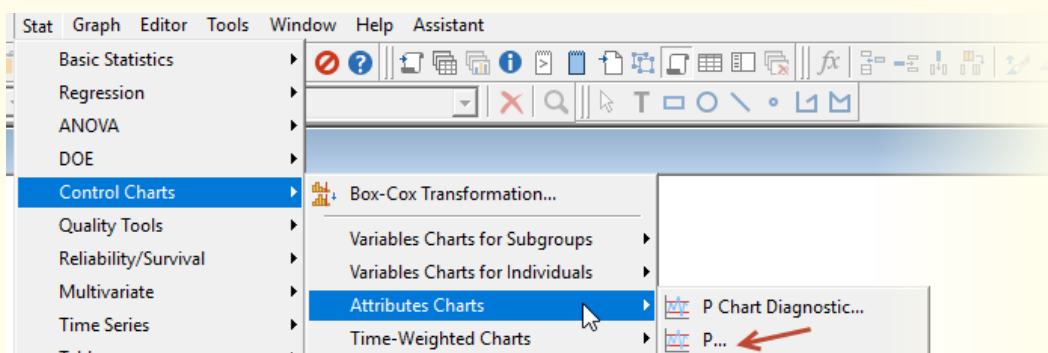


How-To 9.6 (A p chart Minitab 18)

To construct a p chart in Minitab 18, click on *Stat > Control Charts > Attribute Charts > Select P... > Select your data > Input the sample size or the column of sample sizes > OK*. See the snapshot in Figure 9.8.



Figure 9.8: Options for a p chart in Minitab 18



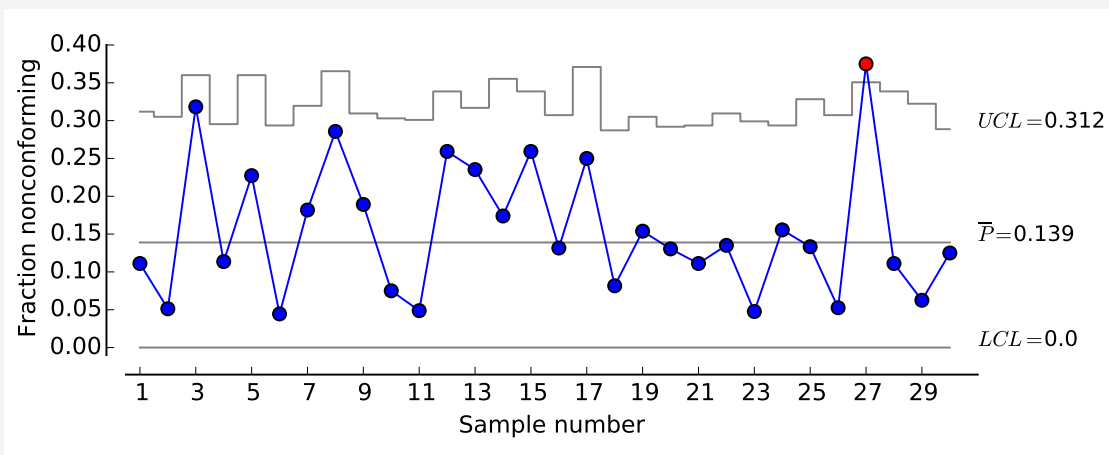
Example 9.5 (A p chart)

The HIM^a manager at Metropolis Hospital sits on the revenue cycle steering committee. At the last week's meeting, it was decided to start monitoring billing denials by the department. The HIM manager was charged with monitoring denials due to medical coding errors. Table 9.2 presents the monthly denial data for the last 30 months. The column of *Denials* represents the variable sample size of all denials, and the *Coding* column contains the number of denials due to medical coding errors.

Table 9.2: Denials related to medical coding errors at Metropolis Hospital

<i>Sample#</i>	<i>Denials</i>	<i>Coding</i>	<i>Sample#</i>	<i>Denials</i>	<i>Coding</i>
1	36	4	16	38	5
2	39	2	17	20	5
3	22	7	18	49	4
4	44	5	19	39	6
5	22	5	20	46	6
6	45	2	21	45	5
7	33	6	22	37	5
8	21	6	23	42	2
9	37	7	24	45	7
10	40	3	25	30	4
11	41	2	26	38	2
12	27	7	27	24	9
13	34	8	28	27	3
14	23	4	29	32	2
15	27	7	30	48	6

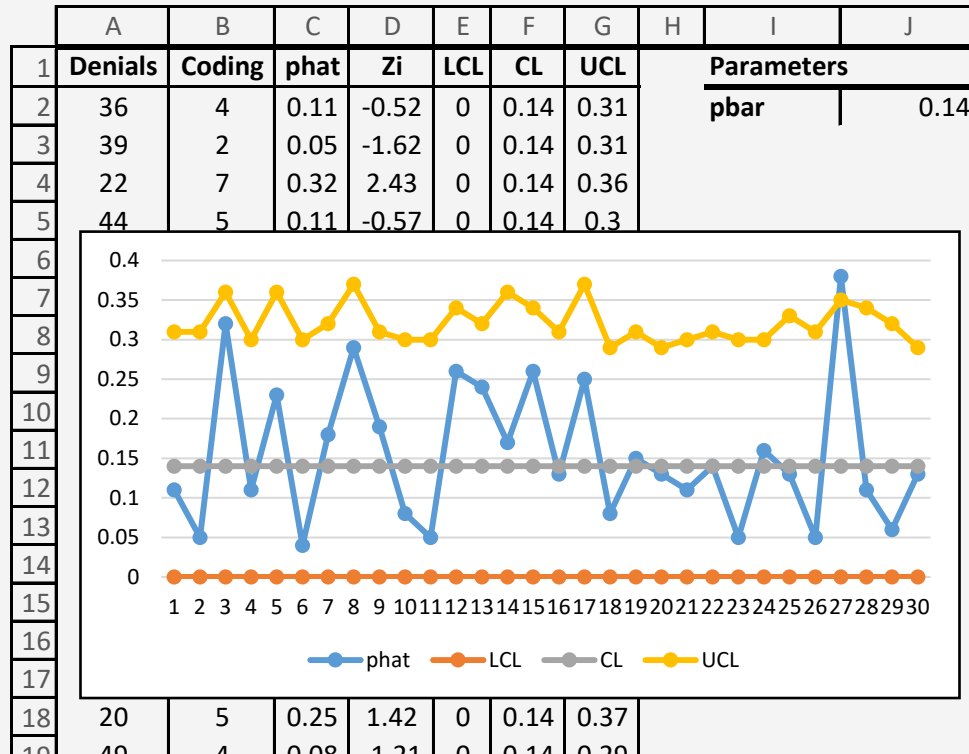
The manager has decided to apply the p chart to monitor the process. Using Python, the manager programmed the formulas in Box 9.3, when p is not given, and produced the chart in Figure 9.9. This p chart exhibits out-of-control behaviors at sample 27. To improve the process, the manager needs to find and fix all the causes of special variation in this sample. Subsequently, sample 27 can be omitted, and a new control chart created to continue to monitor the process. This procedure should be iterated until no more out-of-control behaviors exist in the process.

Figure 9.9: A p chart based on the data in Table 9.2

Next, we demonstrate how you can reproduce this p chart in Excel. We also include instructions about how you can create a corresponding Z chart for standardized values.

1. First set up your spreadsheet, as illustrated in Figure 9.10.

Figure 9.10: A setup of Excel to create the p and Z charts based on the data in Table 9.2



2. Determine the \bar{p} value in cell J2 as indicated in Equation 9.20, as follows:

$$= \text{ROUND}(\text{SUM}(B2 : B31) / \text{SUM}(A2 : A31), 2).$$
3. Calculate the \hat{p}_i statistic for each month i using this ratio: $\text{Coding} / \text{Denials}$. For example, the value in cell C2 was obtained by $= \text{ROUND}(B2 / A2, 2)$. You can drag down this formula to populate the rest of the values.
4. Create control limits when no standards are given, as indicated in Box 9.3. For example, the value of LCL in cell E2 was calculated using $= \text{MAX}(0, \$J\$2 - 3 * \text{SQRT}(\$J\$2 * (1 - \$J\$2) / A2))$. The value of CL in cell F2 was obtained by $= \$J\2 . The value of UCL in cell G2 was determined using $= \text{ROUND}(\$J\$2 + 3 * \text{SQRT}(\$J\$2 * (1 - \$J\$2) / A2), 2)$.
5. To create the p chart, insert line charts for columns C, E, F, and G.
6. You can also create a standardized Z chart. Use Equation 9.29 to calculate the statistic to be monitored, Z_i . For example, the value of Z_i in cell D2 was obtained by $= \text{ROUND}((C2 - \$J\$2) / \text{SQRT}(\$J\$2 * (1 - \$J\$2) / A2), 2)$. You can also drag down this formula to populate the rest of the values.

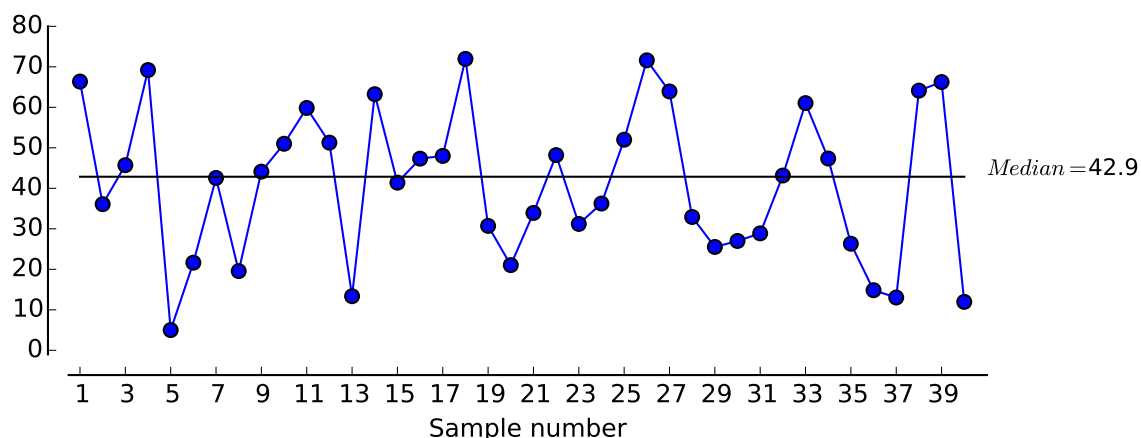
To reproduce the p chart in Minitab, follow the instructions in How-To 9.6.

9.3 Run charts

9.3.1 A general concept

Run charts are like control charts but without limits. Moreover, the centerline in the run charts is typically the median of the data, not the mean. While, whenever possible, we should utilize control charts to monitor processes, using run charts has its advantages such as the simpler setup. Also, in run charts, we do not have to assume any particular probability distribution, meaning that we could use these tools to monitor any process. The major downside of run charts is that the interpretation requires a bit more effort since there is no easy way to tell that a particular sample is out-of-control. Figure 9.11 portrays an example of a run chart that relates to the control chart in Figure 9.3.

Figure 9.11: An example of a run chart based on the chart in Figure 9.2



From the control chart in Figure 9.3, we have previously concluded that there were two areas of special cause variation due to a point falling outside of LCL and a violation of one of the sensitizing rules since 2 out of 3 consecutive points were between the second and third standard deviation limits on the same side. Now the question at hand is that since there are no limits in the run chart in Figure 9.11, how do we go about detecting out-of-control behaviors?

9.3.2 Interpreting run charts

We detect out-of-control behaviors in run charts by finding instances of **non-randomness** in the process. We utilize two techniques to accomplish this task: **visual examination** and statistical **run tests**.

Visual examination

We can use a visual examination to detect special cause variation in run charts by applying the following rules [51]:

1. Seven or more points on the same side of the median signify a likely **shift** in the process.
2. Seven or more points that are either increasing or decreasing indicate a **trend** in the process.

We can also visually detect other likely special causes, such as a **cluster** of points and **zig-zag** or **oscillation** behaviors. To test the significance of these out-of-control behaviors, we apply statistical run tests.

Run tests

We conduct statistical run tests to determine the significance of non-randomness in the process. Our null hypothesis is that the sequence of the process measurements is random. Run tests work for both the **metric** and **non-metric** data. For our purposes, a sequence of groups of data points that are above (**run-ups** +) or below (**run-downs** -) the centerline constitute non-metric data. In contrast, a sequence of successive numerical differences in the process constitutes metric data. We define the term **run** as follows:

Definition 9.1 (Run) A run is a sequence of identical signs (+ or -) until a different sign occurs or the series terminates [57].

To construct a run test, we assume that there are R total number of runs in the data, given by:

$$R = y_1 + y_2 \quad (9.30)$$

where y_1 is the number of run-ups and y_2 is the number of run-downs. We use n_1 to symbolize the total count of points in the set y_1 and n_2 to denote the total count of points in the set y_2 . We calculate $p(y_1, y_2)$, the probability of obtaining exactly y_1 runs and y_2 runs this way [61]:

$$p(y_1, y_2) = \frac{\binom{n_1-1}{y_1-1} \binom{n_2-1}{y_2-1}}{\binom{n_1+n_2}{n_1}} \quad (9.31)$$

We obtain the probability of R being less or equal to some value k , that is $p(R \leq k)$ by adding together the probabilities of different combinations of y_1 and y_2 that produce $y_1 + y_2 \leq k$. We use $p(R \leq k)$ to establish the confidence interval for R . If R falls outside of our interval, we reject the null hypothesis, which implies that out-of-control behaviors exist. For small sample sizes of n_1 and n_2 , we can establish the confidence interval of R by consulting reference tables of run statistics that are often found in standard statistical textbooks (e.g., see Wackerly et al. (2007)[57]). As n_1 and n_2 get larger (e.g., when

both n_1 and n_2 are greater than 10), we assume a normal distribution and establish the confidence interval as follows [57]:

$$E(R) - |Z|\sqrt{V(R)} \leq R \leq E(R) + |Z|\sqrt{V(R)} \quad (9.32)$$

where Z is a standard value given by:

$$Z = \frac{R - E(R)}{\sqrt{V(R)}} \quad (9.33)$$

Here, $E(R)$ symbolizes the expectation of runs, and $V(R)$ represents the variance of runs. When the process generates non-metric data, we obtain $E(R)$ and $V(R)$ this way [61]:

$$E(R) = \frac{2n_1n_2}{n} + 1 \quad (9.34)$$

$$V(R) = \frac{2n_1n_2(2n_1n_2 - n)}{(n-1)n^2} \quad (9.35)$$

where n is the total number of observations given. For metric data, we approximate $E(R)$ and $V(R)$ as follows [61]:

$$E(R) \approx \frac{2n-1}{3} \quad (9.36)$$

$$V(R) \approx \frac{16n-29}{90} \quad (9.37)$$

Interpretation

When $R < E(R) - |Z|\sqrt{V(R)}$, we have **too few runs**, whereas $R > E(R) + |Z|\sqrt{V(R)}$ implies that we have **too many runs**. In general, having too many or too few runs is indicative of non-randomness in the process data [57]. In non-metric data, having **too few runs** signals the tendency of the data to **cluster** above or below the centerline. The result of **too many runs** implies **zig-zag** behaviors, which may mean a **mixture** of two processes. It follows that [42]:

$$(p\text{-value for mixtures}) + (p\text{-value for clustering}) = 1 \quad (9.38)$$

For metric data, having **too few runs** implies the tendency of the data to form **trends** above or below the centerline. Having **too many runs** suggests that **oscillation** behaviors exist in the process. It is given that [42]:

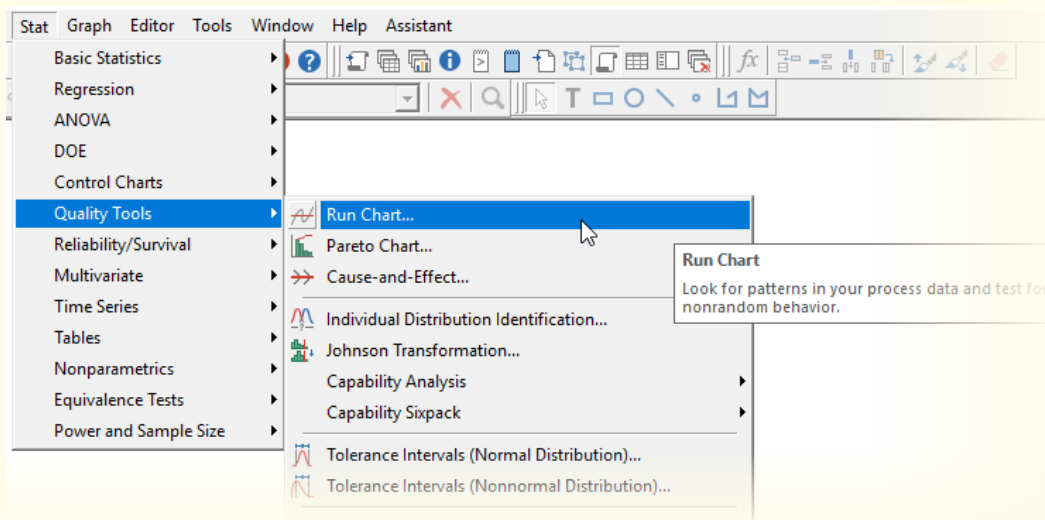
$$(p\text{-value for trends}) + (p\text{-value for oscillation}) = 1 \quad (9.39)$$

How-To 9.7 (Run charts in Minitab 18)

Click on *Stat > Quality Tools > Run Chart > check the single column radio option > select your data > Input the subgroup size > Plot subgroup medians > OK.*



Figure 9.12: The options for run charts in Minitab 18

**How-To 9.8 (Run charts in Excel 2013)**

Excel does not have a built-in option to create run charts, but we can manually program the given formulas as we demonstrate in Examples 9.6 - 9.8.

**Example 9.6 (Binary process data)**

Metropolis Hospital monitors patient satisfaction. On the survey questionnaire, patients are asked whether they are satisfied (Y) or not satisfied (N) with their hospital stay. The following is the sequence of the responses from the most recent ten discharges:

$$Y, Y, N, Y, N, N, Y, Y, N, Y \quad (9.40)$$

The officials at Metropolis Hospital have asked us to help answer the question of whether these responses are random. To answer this question, we begin by enumerating the runs in the data as follows:

$$\underbrace{Y, Y}_{R1}, \underbrace{N}_{R2}, \underbrace{Y}_{R3}, \underbrace{N, N}_{R4}, \underbrace{Y, Y}_{R5}, \underbrace{N}_{R6}, \underbrace{Y}_{R7} \quad (9.41)$$

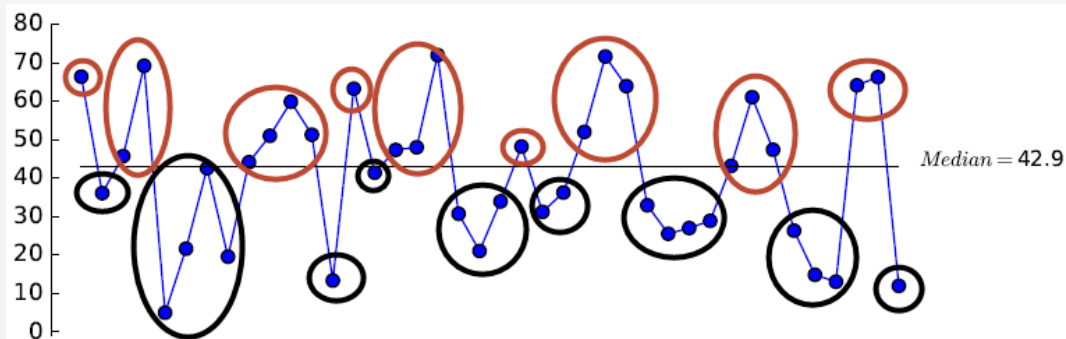
Expression 9.41 indicates that we have 7 runs meaning that $R = 7$. We represent run-downs ($R1, R3, R5$, and $R7$) with y_1 . We notice that the number of data points in y_1 corresponds to $n_1 = 6$. We represent run-ups ($R2, R4$, and $R6$) with y_2 where $n_2 = 4$. By consulting reference tables of runs, when $n_1 = 6$ and $n_2 = 4$, we establish the confidence interval of $p(R \leq 2) = 0.010$ and $p(R > 8) = 1 - p(R \leq 8) = 0.024$.

Given that $R = 7$ falls within this confidence interval, **we fail to reject the null hypothesis** that the sequence of patient responses is random.

Example 9.7 (Non-metric data from run charts)

We revisit the run chart depicted in Figure 9.11 and attempt to identify out-of-control behaviors in this chart. We start by counting the number of runs above and below the median, as we illustrate in Figure 9.13.

Figure 9.13: Counting the number of runs in the chart portrayed in Figure 9.11



From Figure 9.13, we identify 9 run-ups with 20 data points. We represent this information with $y_1 = 9$ and $n_1 = 20$. Coincidentally, the information for run-downs is identical as follows: $y_2 = 9$ and $n_2 = 20$. The total number of runs is given by $R = y_1 + y_2 = 9 + 9 = 18$. We use Equations 9.33 - 9.35 to construct the following confidence interval:

$$E(R) = 21 \quad (9.42)$$

$$V(R) = 9.74 \quad (9.43)$$

$$|Z| = 1.96 \quad (9.44)$$

$$15 \leq R \leq 27 \quad (9.45)$$

We set the significance level to $\alpha = 0.05$. Given that $R = 18$ falls within the resulting confidence interval, **we fail to reject the null hypothesis** that this sequence is random. We compute the corresponding *p-value for clustering* as follows:

$$p\text{-value} \equiv \Theta^{-1}(Z_0) = 0.168 \quad (9.46)$$

where Z_0 is the test statistic obtained this way:

$$Z_0 = \frac{18 - 21}{\sqrt{9.74}} = -0.96108 \quad (9.47)$$

Since $p\text{-value} > 0.05$, we reject the null hypothesis that the process does not have clusters. The *p-value for mixtures* is given by $1 - 0.168 = 0.832$, which leads us to conclude that there is no statistically significant process mixtures in this example.

Example 9.8 (Metric data from run charts)

Let us revisit Example 9.7, but this time investigate the randomness in the sequence of successive numerical differences using the data in Table 9.3.

Table 9.3: Data related to the run chart in Figure 9.13

<i>Sample</i>	<i>Data</i>	<i>Sample</i>	<i>Data</i>
1	66.36	21	33.92
2	36.08	22	48.2
3	45.72	23	31.2
4	69.2	24	36.24
5	5.0	25	52.0
6	21.64	26	71.6
7	42.56	27	63.92
8	19.56	28	32.92
9	44.12	29	25.52
10	51.0	30	27.0
11	59.84	31	28.88
12	51.28	32	43.16
13	13.36	33	61.04
14	63.24	34	47.36
15	41.4	35	26.32
16	47.32	36	14.84
17	48.0	37	13.04
18	71.96	38	64.12
19	30.72	39	66.24
20	21.04	40	11.96

We begin by determining the signs of the running differences, as illustrated in Table 9.4. This table has two derived columns. We use the column of *SIGN* to track the signs of the differences and the column of *COUNT SIGN* to identify runs. For example, we determined the sign in cell B3 using the *SIGN()* function in Excel as follows: $SIGN(36.08 - 66.36) = -$. Likewise, we obtained the sign in cell B4 this way: $SIGN(45.72 - 36.08) = +$. To automate this procedure, we encoded the following formula in cell B3:

$$= IF(SIGN(A3 - A2) = 1, " + ", " - ") \quad (9.48)$$

Next, we dragged down this formula to populate the rest of the signs. In the *COUNT SIGN* column, we encoded 1 if a run existed per Definition 9.1. For example, the value of 1 in cell C3 was obtained using $IF(B4 = B5, 0, 1)$. Dragging down this formula allowed us to populate the rest of the values. The sum of the *COUNT SIGN* column gave us the total number of runs R . To find the number in the run-ups (+) or n_1 , we used this formula:

$$= COUNTIF(B3 : B41, " + ") \quad (9.49)$$

Correspondingly, we found the number of the run-downs(-) or n_2 like this:

$$= COUNTIF(B3 : B41, " - ") \quad (9.50)$$

In the end, we determined that $R = 19$, $n_1 = 22$, and $n_2 = 17$.

Table 9.4: A spreadsheet setup for creating run tests of the data in Table 9.3

	A	B	C
1	Data	SIGN	COUNT SIGN
2	66.36		
3	36.08	-	1
4	45.72	+	0
5	69.2	+	1
6	5.0	-	1
7	21.64	+	0
8	42.56	+	1
9	19.56	-	1
10	44.12	+	0
11	51.0	+	1
...

Finally, we applied the formulas in Equations 9.33, 9.36, and 9.37 to find:

$$E(R) = 26.3 \quad (9.51)$$

$$V(R) = 6.78 \quad (9.52)$$

$$|Z| = 1.96 \quad (9.53)$$

$$21.2 \leq R \leq 31.4 \quad (9.54)$$

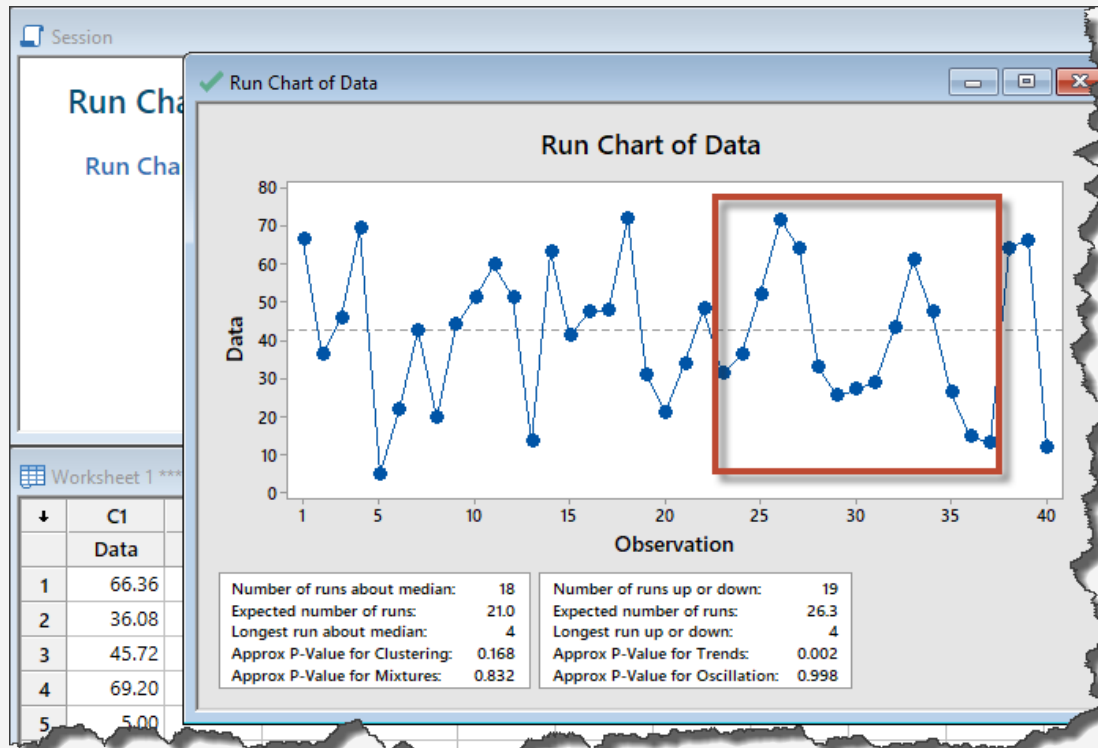
Since $R = 19$ falls outside of the confidence interval, **we reject the null hypothesis** at the significance level of 0.05. Additionally, since $R < E(R) - |Z|\sqrt{V(R)}$, that is $19 < 21.2$, we conclude that a **trend** is likely in the process. To estimate the *p-value* for the trend, we proceed this way:

$$Z_0 = \frac{19 - 26.3}{\sqrt{6.78}} = -2.814 \quad (9.55)$$

We obtain the probability of interest this way:

$$p\text{-value} \equiv \Theta^{-1}(Z_0) = 0.002 \quad (9.56)$$

Since $p\text{-value} < 0.05$, we reject the null hypothesis that there is no trend in the data. We obtain the *p-value* for oscillation by $1 - 0.002 = 0.998$. Since $p\text{-value} > 0.05$, we conclude that no statistically significant oscillation exists in this process. By following instructions in How-To 9.7, we reproduced similar results in Minitab, as portrayed in Figure 9.14. We used a subgroup of size 1.

Figure 9.14: The run chart in Minitab based on the data in Table 9.3

The boxed area in Figure 9.14 shows sequences that exhibit trend-like behaviors. As noted earlier, unlike control charts, run charts don't allow us to pinpoint special cause variation to a particular sample number.

9.4 Out-of-control action plan (OCAP)

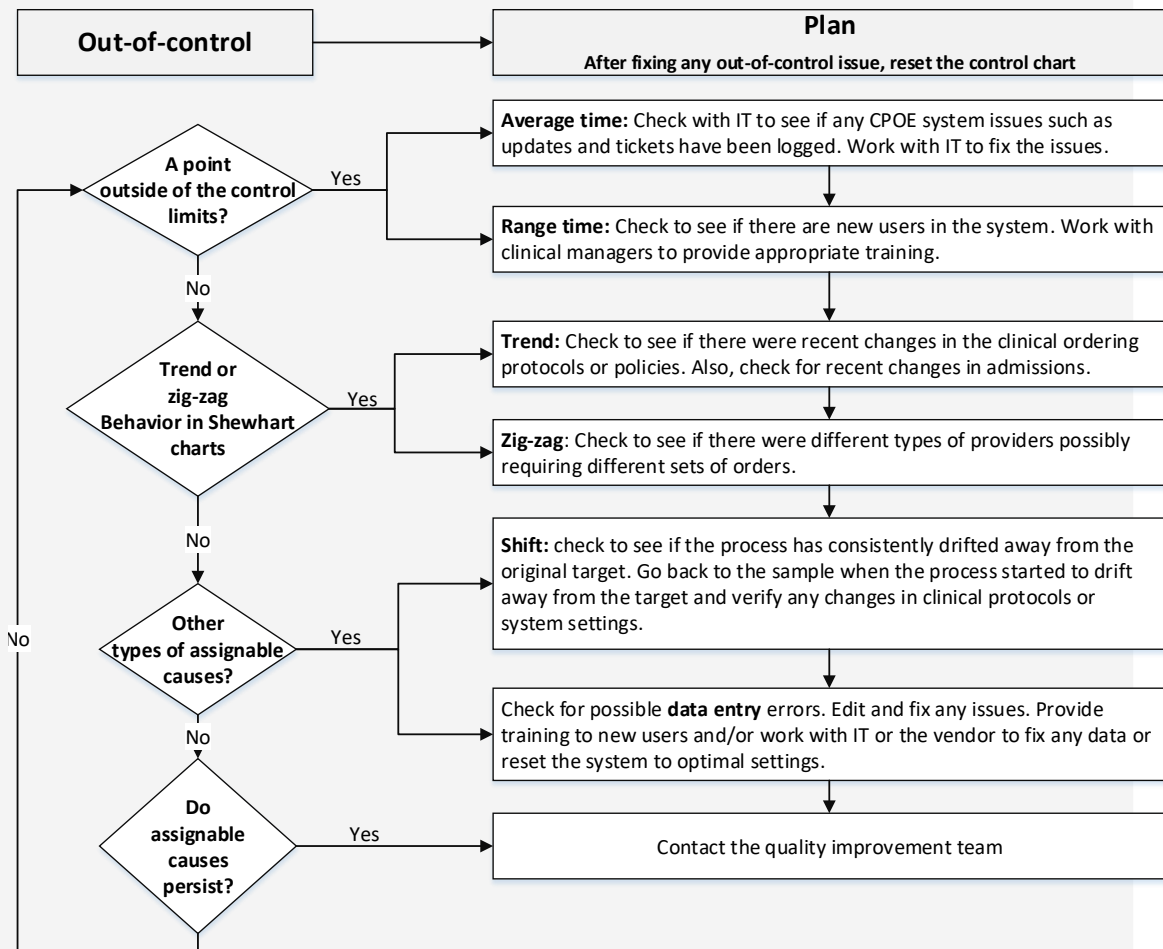
Once we have improved and stabilized the process, we move to the *Control* period of DMAIC, where we implement phase II control charts to monitor small-shifts in process. But, no process stays stable forever. Over time, out-of-control behaviors will arise again, even when our improvement safeguards remain in place. We must not only be able to monitor and detect new out-of-control behaviors but also must have a plan of action to address any rising new problems and false-alarms. We refer to such a plan as OCAP and implement it during the *Control* period of DMAIC. We typically organize an OCAP document as a flow chart with two primary components: **checkpoints**, which are decision nodes for detecting special cause variation, and **terminators**, which are plausible corrective actions [43]. We illustrate with a generic example in Figure 9.15.

Example 9.9 (OCAP)

You are a member of a quality team at Central City Hospital. This hospital has just finished improving the order entry process using a computerized order entry

(CPOE) system. Your team is now working on the control phase of DMAIC by writing an OCAP document for monitoring the daily mean response of orders in the process. Your final document is shown in Figure 9.15. The owner of the new process will use this document to detect and attempt to correct out-of-control behaviors in the *ImR* chart.

Figure 9.15: A generic OCAP document about a CPOE process at Central City Hospital

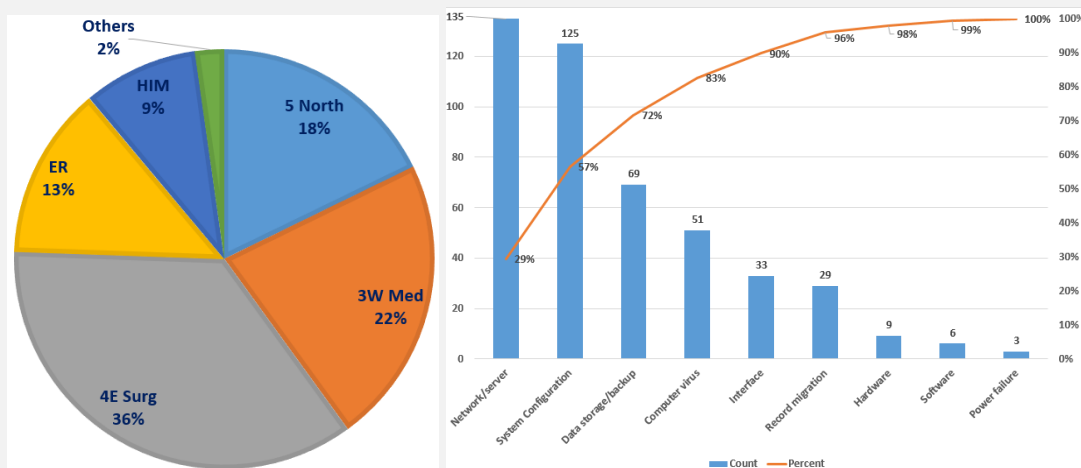


9.5 Dashboards

Besides the OCAP document, we also must have a mechanism for visualizing the process performance. One of the tools that apply for this purpose is a dashboard. The latter is a collection of data visualization tools that we assemble to display a few key indicators of

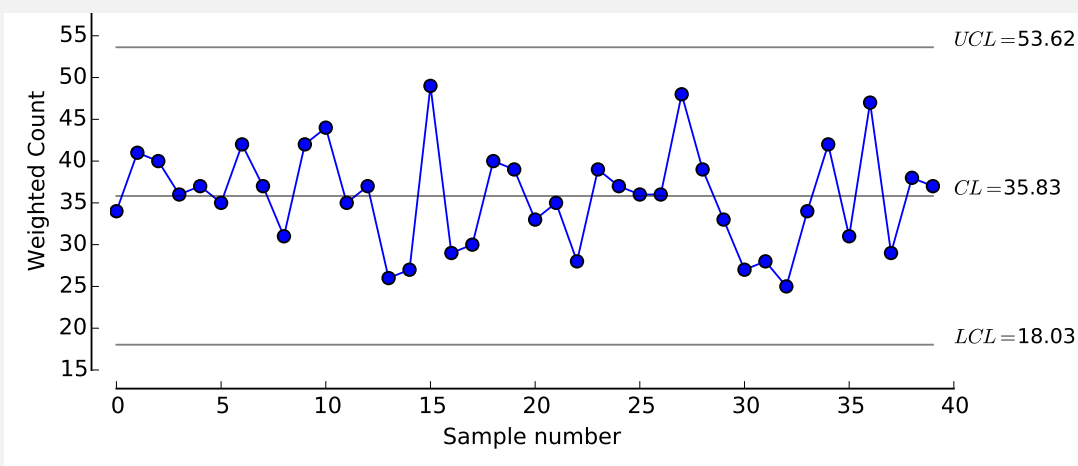
the process graphically. When selecting visuals to put on a dashboard, we must ensure that the intended audience can interpret the presented information. Depending on the software we use, we should also give the owner an option to drill down the data to gain insights into the performance at the basic unit of the process [60]. We show an example of a dashboard in Box 9.4. All graphics shown there can easily be created using Excel or specialized software such as Minitab or QIMacros [54].

Box 9.4 An informatics process, Central City Hospital, January 2019



(a) Quality issue by source: Jan 2019

(b) Quality issue by type: Jan, 2019



(c) Monthly control chart of the weighted number of IT events up to Jan 2019

9.6 Transitioning the process back to the owner

If we were an outside team working on the quality improvement project, we would have to hand the process back to the owner during the *Control* phase of DMAIC. As part of the transition, the owner must agree to the OCAP of the new process [24]. Furthermore, we must create a report for the owner detailing all that we have accomplished, including a record of any quality gains and any cost savings that we have achieved. This kind of documentation is particularly important for executives who make decisions about funding quality improvement projects. Finally, we must create tools (e.g., dashboards) to allow all stakeholders to be part of monitoring the performance of the new process.

9.7 Check Your Understanding

1. If we want to monitor the average coding time per record, what type of control charts are we likely to use?
 - (a) Variable control charts
 - (b) Attribute control charts
2. Given individual observations, which of the following control charts are we likely to use to monitor variability?
 - (a) \overline{MR}
 - (b) \overline{R}
 - (c) \overline{S}
 - (d) S^2
3. Removing ——— results in the creation of a new process.
 - (a) special cause variation from an existing process
 - (b) common cause variation from an existing process
4. When there are at least eight consecutive points on the same side of the centerline of a control chart, we likely have ———
 - (a) a shift in the process.
 - (b) zig-zag behaviors in the process.
 - (c) trend behaviors in the process.
 - (d) oscillation behaviors in the process.
5. ARL_α in a stable and normally distributed process with $L = 2$ is about:
 - (a) 3
 - (b) 10
 - (c) 20
 - (d) 370
6. The g chart is related to which random variable?
 - (a) Geometric
 - (b) Poisson
 - (c) Binomial
 - (d) Normal

7. From the operating curves, we are likely to observe that as the sample size increases,
- (a) α decreases
 - (b) β decreases
 - (c) α increases
 - (d) β increases
8. How are we likely to monitor a process with variable data for individual observations?
- (a) Using ImR charts
 - (b) Using $XbarR$ charts
 - (c) Using $XbarS$ charts
 - (d) Using p charts
9. The ImR chart is related to which random variable?
- (a) Geometric
 - (b) Poisson
 - (c) Binomial
 - (d) Normal
10. Your hypothesis is as follows:
- $$H_0 : \text{the process in control}$$
- $$H_1 : \text{the process not in control}$$
- You noticed that a point fell outside of the control limits and concluded that a special cause variation existed. Your conclusion was equivalent to:
- (a) rejecting the null hypothesis.
 - (b) failing to reject the null hypothesis.
11. Sensitizing rules are mainly applied to detect ——
- (a) large shifts in the process.
 - (b) small shifts in the process.
12. The person credited with introducing control charts is ——
- (a) Deming.
 - (b) Juran.
 - (c) Ishikawa.

- (d) Shewhart.
13. ARL_β is also referred to as —
- (a) in-control ARL .
 - (b) out-of-control ARL .
14. In a p chart with $L = 3$ and $\bar{p} = 0.5$, what is the minimum number of samples to collect to ensure that $LCL > 0$?
- (a) 7
 - (b) 8
 - (c) 9
 - (d) 10
15. To monitor the rate of critical alert overrides in a clinical system, we are likely to use which of the following control charts?
- (a) c
 - (b) g
 - (c) p
 - (d) u
16. To monitor the number of *surgeries on the wrong patients* in a typical hospital, we are more likely to use which set of the following control charts?
- (a) p and np
 - (b) c and u
 - (c) g and h
 - (d) x and s
17. The c chart is related to which random variable?
- (a) Geometric
 - (b) Poisson
 - (c) Binomial
 - (d) Normal
18. Too many runs in a *run* chart of non-metric data indicate the tendency of the process to —
- (a) zig-zag.
 - (b) oscillate.
 - (c) cluster.

- (d) trend.
19. A trend in a *run* chart is defined as — or more points that are either increasing or decreasing.
- (a) five
 - (b) seven
 - (c) fourteen
 - (d) twenty-one
20. After analyzing the sequence of the differences in the metric data of a *run* chart, you found too many runs. This likely indicates the presence of — in the process.
- (a) a trend
 - (b) an oscillation
 - (c) a cluster
 - (d) a shift
21. You were creating a *run* chart using non-metric data and found that n_1 for run-ups was 14, and n_2 for run-downs was 14. What was your $V(R)$?
- (a) 1.96
 - (b) 6.74
 - (c) 11.03
 - (d) 14.00
22. To detect small shifts in a process, it is recommended to use —
- (a) Shewhart control charts.
 - (b) time-weighted control charts.
 - (c) run charts.
 - (d) operating-characteristic curves.
23. Table 9.5 presents data about the *time to administration of pain medication to emergency department (ED) patients with broken bones* at Metropolis Hospital. Tracking and reporting this information is required by CMS as one of its mandated quality measures about timely and effective care [17].
- (a) Using Excel and Minitab, create a run chart
 - (b) Use the same dataset to create *ImR* charts.
 - (c) Compare and contrast your results. What can you conclude?

Table 9.5: Time to administration of pain medication to ED patients with broken bones at Metropolis Hospital

<i>Patient#</i>	<i>Angle</i>	<i>Patient#</i>	<i>Angle</i>
1	42	21	40
2	46	22	47
3	17	23	47
4	25	24	9
5	41	25	44
6	38	26	24
7	23	27	8
8	55	28	23
9	28	29	36
10	12	30	55
11	14	31	63
12	26	32	60
13	28	33	53
14	32	34	26
15	11	35	26
16	53	36	13
17	20	37	11
18	5	38	8
19	8	39	29
20	15	40	75

24. The manager of the ED at Metropolis Hospital would like to monitor and control the number of patients leaving without being seen (LWBS). Table 9.6 presents samples collected on weekdays. Implement appropriate control charts to monitor the rate and the number of LWBS. If any point falls outside of the limits, remove it, and recreate the charts. Repeat this procedure until the process is stable.

(a) What can you conclude?

(b) How would this process perform against the LWBS target of 2.5%?

Table 9.6: Weekday samples of LWBS in the ED at Metropolis Hospital, 2018

<i>Sample#</i>	<i>Visits</i>	<i>LWBS</i>	<i>Sample#</i>	<i>Visits</i>	<i>LWBS</i>
1	221	7	31	182	5
2	217	10	32	245	3
3	201	9	33	197	8
4	200	6	34	239	7
5	247	4	35	204	6
6	231	7	36	213	4
7	243	5	37	180	5
8	240	7	38	208	11
9	246	7	39	231	7
10	225	4	40	195	6
11	247	9	41	201	7
12	202	4	42	230	4
13	233	7	43	203	7
14	216	8	44	236	8
15	215	5	45	191	6
16	224	4	46	230	9
17	220	6	47	234	8
18	201	4	48	229	3
19	198	8	49	231	9
20	225	8	50	224	3
21	250	7	51	226	4
22	245	9	52	201	3
23	225	9	53	249	7
24	225	8	54	225	8
25	209	7	55	181	8
26	237	8	56	216	7
27	181	7	57	220	6
28	196	6	58	249	16
29	183	6	59	216	4
30	233	9	60	184	6

25. Table 9.7 presents the frequency of the monthly overrides of critical alerts in the CPOE system at Metro City Hospital.

(a) Using Excel, create the corresponding run chart. What out-of-control behaviors can you visually detect?

- (b) Calculate run statistics and appropriate confidence intervals in Excel using $\alpha = 0.05$. What can you conclude about the hypothesis of the process being stable? Explain.
- (c) Create an appropriate control chart.
- (d) Compare and contrast the results from your control chart and run charts. What can you conclude?

Table 9.7: Critical alert overrides in the CPOE system at Metro City Hospital, 2015-2019

Month	#Overrides	Month	#Overrides
1	42	21	40
2	46	22	47
3	17	23	47
4	25	24	9
5	41	25	44
6	38	26	24
7	23	27	8
8	55	28	23
9	28	29	36
10	12	30	55
11	14	31	63
12	26	32	60
13	28	33	53
14	32	34	26
15	11	35	26
16	53	36	13
17	20	37	11
18	5	38	8
19	8	39	29
20	15	40	75

MODULE 10

Appendix

.1 Appendix

Table 1: Factors for Constructing Variable Control Charts

Chart for Averages				Chart for standard deviations				Chart for Ranges							
	Factors for Control Limits			Factors for:				Factors for CL		Factors for Control Limits					
				CL	Control Limits										
<i>n</i>	<i>A</i>	<i>A</i> ₂	<i>A</i> ₃	<i>c</i> ₄	<i>B</i> ₃	<i>B</i> ₄	<i>B</i> ₅	<i>B</i> ₆	<i>d</i> ₂	1/ <i>d</i> ₂	<i>d</i> ₃	<i>D</i> ₁	<i>D</i> ₂	<i>D</i> ₃	<i>D</i> ₄
2	2.121	1.880	2.659	0.7979	0	3.267	0	2.606	1.128	0.8862	0.852	0	3.686	0	3.266
3	1.732	1.023	1.954	0.8862	0	2.568	0	2.276	1.693	0.5908	0.888	0	4.357	0	2.574
4	1.500	0.729	1.628	0.9213	0	2.266	0	2.088	2.059	0.4857	0.879	0	4.697	0	2.281
5	1.342	0.577	1.427	0.9400	0	2.089	0	1.964	2.326	0.4299	0.864	0	4.918	0	2.114
6	1.225	0.483	1.287	0.9515	0.030	1.970	0.029	1.874	2.534	0.3946	0.848	0	5.078	0	2.003
7	1.134	0.419	1.182	0.9594	0.118	1.882	0.113	1.806	2.704	0.3698	0.833	0.206	5.203	0.076	1.924
8	1.061	0.373	1.099	0.9650	0.185	1.815	0.179	1.751	2.847	0.3512	0.819	0.389	5.306	0.137	1.863
9	1.000	0.337	1.032	0.9693	0.239	1.761	0.232	1.707	2.970	0.3367	0.807	0.548	5.392	0.184	1.816
10	0.949	0.308	0.975	0.9727	0.284	1.716	0.276	1.669	3.078	0.3249	0.797	0.688	5.467	0.223	1.777
11	0.905	0.285	0.927	0.9754	0.321	1.679	0.313	1.637	3.173	0.3152	0.787	0.813	5.533	0.256	1.744
12	0.866	0.266	0.886	0.9776	0.354	1.646	0.346	1.610	3.258	0.3069	0.778	0.924	5.593	0.284	1.716
13	0.832	0.249	0.850	0.9794	0.382	1.618	0.374	1.585	3.336	0.2998	0.770	1.026	5.646	0.307	1.693
14	0.802	0.235	0.817	0.9810	0.406	1.594	0.399	1.563	3.407	0.2935	0.763	1.119	5.695	0.328	1.672
15	0.775	0.223	0.789	0.9823	0.428	1.572	0.421	1.544	3.472	0.2880	0.756	1.204	5.739	0.347	1.653
16	0.750	0.212	0.763	0.9835	0.448	1.552	0.440	1.526	3.532	0.2831	0.750	1.283	5.781	0.363	1.637
17	0.728	0.203	0.739	0.9845	0.466	1.534	0.458	1.511	3.588	0.2787	0.744	1.357	5.819	0.378	1.622
18	0.707	0.194	0.718	0.9854	0.482	1.518	0.475	1.496	3.640	0.2747	0.738	1.425	5.855	0.392	1.608
19	0.688	0.187	0.698	0.9862	0.497	1.503	0.490	1.483	3.689	0.2711	0.733	1.490	5.888	0.404	1.596
20	0.671	0.180	0.680	0.9869	0.510	1.490	0.504	1.470	3.735	0.2677	0.728	1.550	5.920	0.415	1.585
21	0.655	0.173	0.663	0.9876	0.523	1.477	0.516	1.459	3.778	0.2647	0.724	1.607	5.950	0.425	1.575
22	0.640	0.167	0.647	0.9882	0.534	1.466	0.528	1.448	3.819	0.2618	0.719	1.661	5.978	0.435	1.565
23	0.626	0.162	0.633	0.9887	0.545	1.455	0.539	1.438	3.858	0.2592	0.715	1.712	6.004	0.444	1.556
24	0.612	0.157	0.619	0.9892	0.555	1.445	0.549	1.429	3.895	0.2567	0.712	1.761	6.030	0.452	1.548
25	0.600	0.153	0.606	0.9896	0.565	1.435	0.559	1.420	3.931	0.2544	0.708	1.807	6.055	0.460	1.540

For $n > 25$:

$$\begin{aligned}
 A &= \frac{3}{\sqrt{n}} \\
 c_4 &\approx \frac{4(n-1)}{4n-3} \\
 A_3 &= \frac{A}{c_4} \\
 B_3 &= 1 - \frac{3}{c_4 \sqrt{2(n-1)}} \\
 B_4 &= 1 + \frac{3}{c_4 \sqrt{2(n-1)}} \\
 B_5 &= c_4 - \frac{3}{2(n-1)} \\
 B_6 &= c_4 + \frac{3}{2(n-1)}
 \end{aligned}$$

For $n \leq 25$ see Table 1.

Bibliography

- [1] Minitab 18. Methods and formulas for binomial data in analysis of means. <https://support.minitab.com/en-us/minitab/18/help-and-how-to/modeling-statistics/anova/how-to/analysis-of-means/methods-and-formulas/binomial-data/>, 2018. [Online; accessed 4-November-2018].
- [2] Hervé Abdi. The kendall rank correlation coefficient. *Encyclopedia of Measurement and Statistics*. Sage, Thousand Oaks, CA, pages 508–510, 2007.
- [3] AHRQ. Mistake-proofing the design of health care processes. <https://archive.ahrq.gov/professionals/quality-patient-safety/patient-safety-resources/resources/mistakeproof/index.html>, 2017.
- [4] Farrokh Alemi and Douglas W Oliver. Tutorial on risk-adjusted p-charts. *Quality Management in Healthcare*, 10(1):1–9, 2001.
- [5] AMIA. Clinical informatics. <https://www.amia.org/applications-informatics/clinical-informatics>, 2017. [Online; accessed 2- September-2017].
- [6] AMIA. Health informatics accreditation standards. https://www.cahiim.org/docs/default-source/about-cahiim/history/final-amia-health-informatics-core-competencies-for-cahiim.pdf?sfvrsn=fed062b8_4, 2017.
- [7] James C Benneyan. Statistical quality control methods in infection control and hospital epidemiology, part i introduction and basic theory. *Infection Control & Hospital Epidemiology*, 19(3):194–214, 1998.
- [8] Tim Benson. *Principles of health interoperability HL7 and SNOMED*. Springer London:, 2010.

- [9] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.
- [10] Åke Blomqvist. The doctor as double agent: Information asymmetry, health insurance, and medical care. *Journal of Health Economics*, 10(4):411–432, 1991.
- [11] Åke Blomqvist and Pierre Thomas Léger. Information asymmetry, insurance, and the decision to hospitalize. *Journal of health economics*, 24(4):775–793, 2005.
- [12] Felicia M Bowens, Patricia A Frye, and Warren A Jones. Health information technology: integration of clinical workflow into meaningful use of electronic health records. *Perspectives in Health Information Management/AHIMA, American Health Information Management Association*, 7(Fall), 2010.
- [13] Travis Breau and Annie Antón. Analyzing regulatory rules for privacy and security requirements. *IEEE transactions on software engineering*, 34(1):5–20, 2008.
- [14] Forrest W Breyfogle III. *Implementing six sigma: smarter solutions using statistical methods*. John Wiley & Sons, 2003.
- [15] Mary Butler. Preventing healthcare’s top four documentation disasters. *Journal of AHIMA*, 86(7):18–23, 2015.
- [16] CMS. nital antibiotic selection for community-acquired pneumonia (cap) in immuno-competent patients. <https://ecqi.healthit.gov/ecqm/measures/cms188v6>, 2018. [Online; accessed 22-December-2018].
- [17] CMS. Quality measures. <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityMeasures/index.html>, 2018. [Online; accessed 4-November-2018].
- [18] Joint Commision. Measures. https://www.jointcommission.org/core_measure_sets.aspx, 2018. [Online; accessed 4-November-2018].
- [19] Lawrence T DeCarlo. On the meaning and use of kurtosis. *Psychological methods*, 2(3):292, 1997.
- [20] Avedis Donabedian. The quality of care: how can it be assessed? *Jama*, 260(12):1743–1748, 1988.
- [21] Molla S Donaldson, Janet M Corrigan, Linda T Kohn, et al. *To err is human: building a safer health system*, volume 6. National Academies Press, 2000.
- [22] C.L Gapenski. *Fundamentals of Healthcare Finance, Second Edition*. Health Administration Press, 2012.
- [23] David Garvin. Competing on the eight dimensions of quality. *Harvard Business Review*, pages 101–109, 1987.
- [24] Michael L George. *Lean six sigma for service*. McGraw-Hill New York, NY, 2003.

- [25] Thomas R Gillem. Deming's 14 points and hospital quality: Responding to the consumer's demand for the best value health care. *Journal of Nursing Care Quality*, 2(3):70–78, 1988.
- [26] The Leafrog Group. Hospital ratings and reports. <http://www.leapfroggroup.org/ratings-reports>, 2018. [Online; accessed 4-November-2018].
- [27] Randolph W Hall. *Queueing methods: for services and manufacturing*. Pearson College Div, 1991.
- [28] Jeffrey A Hoffer, Venkataraman Ramesh, and Heikki Topi. *Modern database management*. Pearson, 2016.
- [29] Brent C James, Mark R Chassin, Harold I Goldberg, John G King, and James S Todd. Implementing practice guidelines through clinical quality improvement; improving quality of care with practice guidelines; should we be implementing untested guidelines?; the relevance of practical experience to american hospitals; quest for quality or cost containment; reply. *Frontiers of Health Services Management*, 10(1):3, 1993.
- [30] Otis B Jennings, Avishai Mandelbaum, William A Massey, and Ward Whitt. Server staffing to meet time-varying demand. *Management Science*, 42(10):1383–1394, 1996.
- [31] Charles Higgins Kepner and Benjamin B Tregoe. *The rational manager: A systematic approach to problem solving and decision making*. McGraw-Hill, 1965.
- [32] Maríosa Kieran, Mary Cleary, Aoife De Brún, and Aileen Igoe. Supply and demand: application of lean six sigma methods to improve drug round efficiency and release nursing time. *International Journal for Quality in Health Care*, 29(6):803–809, 2017.
- [33] Mi Ok Kim, Enrico Coiera, and Farah Magrabi. Problems with health information technology and their effects on care delivery and patient outcomes: a systematic review. *Journal of the American Medical Informatics Association*, 24(2):246–250, 2017.
- [34] Sameer Kumar and Krista Aldrich. Overcoming barriers to electronic medical record (emr) implementation in the us healthcare system: A comparative study. *Health informatics journal*, 16(4):306–318, 2010.
- [35] Russell V Lenth. Quick and easy analysis of unreplicated factorials. *Technometrics*, 31(4):469–473, 1989.
- [36] Douglas Lind, William Marchal, and Samuel Wathen. *Basic Statistics for Business and Economics with Student CD*. McGraw-Hill/Irwin, 2007.
- [37] Douglas A Lind, William G Marchal, and M Wathen. *Statistical Techniques in Business & Economics*. USA, McGraw-Hill/Irwin, 2007.

- [38] John DC Little. A proof for the queuing formula: $L = \lambda w$. *Operations Research*, 9(3):383–387, 1961.
- [39] Youssra Marjoua and Kevin J Bozic. Brief history of quality movement in us health-care. *Current reviews in musculoskeletal medicine*, 5(4):265–273, 2012.
- [40] Andrew McAfee and Erik Brynjolfsson. *Machine, platform, crowd: Harnessing our digital future*. WW Norton & Company, 2017.
- [41] Medicare.gov. Hospital compare. <https://www.medicare.gov/hospitalcompare/search.html?>, 2018. [Online; accessed 4-November-2018].
- [42] Minitab. Methods and formulas for run chart. <https://support.minitab.com/en-us/minitab/18/help-and-how-to/quality-and-process-improvement/quality-tools/how-to/run-chart/methods-and-formulas/methods-and-formulas/>, 2018. [Online; accessed 4-November-2018].
- [43] Douglas C Montgomery. *Introduction to statistical quality control*. John Wiley & Sons, 2007.
- [44] Henry R Neave. Deming's 14 points for management: framework for success. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 36(5):561–570, 1987.
- [45] Henry R Neave. *The deming dimension*. SPC Press Knoxville, TN, 1990.
- [46] Jerome Niyirora and Jamol Pender. Optimal staffing in nonstationary service centers with constraints. *Naval Research Logistics (NRL)*, 63(8):615–630, 2016.
- [47] Jerome Niyirora and Jun Zhuang. Fluid approximations and control of queues in emergency departments. *European Journal of Operational Research*, 261(3):1110–1124, 2017.
- [48] NYDOH. Nys hospital profiles. <https://profiles.health.ny.gov/hospital>, 2018. [Online; accessed 4-November-2018].
- [49] Pamela K Oachs and Amy L Watters. *Health information management: Concepts, principles, and practice*. AHIMA, 2016.
- [50] Gregory S Ogrinc. *Fundamentals of Health Care Improvement: A Guide to Improving Your Patient's Care*. Joint Commission Resources, 2012.
- [51] Gregory S Ogrinc et al. *Fundamentals of Health Care Improvement: A Guide to Improving Your Patient's Care*. Joint Commission Resources, 2012.
- [52] Ananthanarayanan Parasuraman, Valarie A Zeithaml, and Leonard L Berry. Servqual: A multiple-item scale for measuring consumer perc. *Journal of retailing*, 64(1):12, 1988.

- [53] Thomas Pyzdek. *The Six Sigma handbook: The complete guide for greenbelts, blackbelts, and managers at all levels, revised and expanded edition*. McGraw-Hill New York, 2003.
- [54] QIMacros. Spc software for excel. <https://www.qimacros.com/spc-software-for-excel/>, 2018. [Online; accessed 30-December-2018].
- [55] Thomas K Ross. *Health Care Quality Management: Tools and Applications*. John Wiley & Sons, 2014.
- [56] David J Spiegelhalter. Funnel plots for comparing institutional performance. *Statistics in medicine*, 24(8):1185–1202, 2005.
- [57] Dennis Wackerly, William Mendenhall, and Richard Scheaffer. *Mathematical statistics with applications*. Nelson Education, 2007.
- [58] M Waskom. Seaborn.swarmplot. <https://seaborn.pydata.org/generated/seaborn.swarmplot.html>, 2019. [Online; accessed 3-January-2019].
- [59] Nicole Gray Weiskopf and Chunhua Weng. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *Journal of the American Medical Informatics Association*, 20(1):144–151, 2013.
- [60] Susan E White. *A practical approach to analyzing healthcare data*. American Health Information Management Association, 2016.
- [61] Per Winkel and Nien Fan Zhang. *Statistical development of quality in medicine*, volume 62. John Wiley & Sons, 2007.

- C_p , 84, 150
- C_{pk} , 84
- C_{pl} , 84
- C_{pu} , 84
- ImR* control charts, 199
- R^2 , 104
- $R^2_{adjusted}$, 104
- p control chart, 203
- 14 points, *see* Deming
- 5S+1, 184
- 5Whys diagrams, 117
- 80/20 rule, 19

- Actual capability, 85
- Affinity grouping, 115
- AHRQ, 186
- Analysis of the means, *see* ANOM
- ANOM, 160
- ANOVA, 70, 105, 138, 149, 173, 179, 188
 - One-way ANOVA, 138
 - Two-way ANOVA, 140
- ARL, 195
- Assignable cause, 197
- Average run length, *see* ARL

- Bar charts, 52
- Baseline measures, 28, 46
- Batch processing, 183, 184
- Benchmarking quality, 127
- Box plots, 82
- Brainstorming, 114

- Capability
 - Improvement, 85
- Capability analysis, 84
- Cause-and-effect diagrams, 19, 115
- Central limit theorem, 68
- Checksheets, 19, 48
- Chi-square test, 154, 156
- Coefficient of determination, 104
- Common cause variation, 197
- Confidence interval, 127
- Contingency table, 155
- Contrasts, 173
- Control charts, 20, 186, 193
 - ImR*, 199
 - p , 203
 - Attribute, 199
 - Choosing L , 195
 - Variable, 199
- Correlation, 98
- Covariance, 101
- CPOE, 186
- Crowdsourcing, 114
- Cycle time, 29, 183

- Dashboard, 217
- Data collection, 47
- Data quality, 14, 185
- Database, 48
- Defect, 29
- Defect concentration diagram, 19, 57
- Defective parts per million, *see* DPMO
- Degrees of freedom, 64
- Delphi, 114
- Deming, 20

- Designed Experiments, *see* DOE
- Designing for Six-Sigma, *see* DFSS
- DFSS, 27, 35
- Diagnosis-Related Group, 33
- DMAIC, 17
- DOE, 18, 98, 141, 171
- Donabedian, 13
- DPMO, 19, 28, 31, 85
- Drift, 32

- EHR, 28
- Electronic Health Record, *see* EHR

- F-test, 138
- Factorial design, *see also* DOE
 - 2^2 , 172
 - 2^3 , 174
 - Blocking, 176
 - Center points, 176
 - Fractional design, 174
 - No significant factor, 176
- Factors for Constructing Variable Charts, 228
- Failure Modes and Effects Analysis , *see* FMEA120
- False alarm, 132
- Feigenbaum, 22
- First normal form, 47
- First-pass yield, 46
- Flowcharts, 35
- FMEA, 120
- Force-field analysis, 122
- Funnel charts, 160

- Goodness-of-fit, 103

- Health care informatics, 14
- HIPAA, 27, 185
- Histograms, 19, 75
- HL7, 27
- Human error, 15
- Hypothesis testing, 130

- Interquartile range, *see* IQR82
- Interruptions, 183, 184
- IQR, 82
- Is-Is Not Matrix, 118
- Ishikawa, 22
- Ishikawa diagrams, *see* Cause-and-effect diagrams115

- Juran, 22

- KPIV, 18
- KPOV, 18

- Lean, 29, 183
- Learning curves, 183
- Line charts, 53
- Little's law, 183

- Marginal plots, 98
- Motorola, 32
- Muda, 29
- Multivoting Technique, 119

- Nightingale, 23
- Nominal Group Technique, 119
- Non-value-added, *see* NVA
- NVA, 29, 31, 36, 38

- OCAP, 216
- Offloading, 184
- Operating-characteristic curves, 132
- Out-of-control plan, *see* OCAP19
- Overprocessing, 183, 184
- Overproduction, 183, 185

- P-value, 106, 131
- Paired t-test, 134
- Pareto charts, 19, 53
- PCE, 18, 28, 30
- PCR, 84–86, 89
- PDCA, 17
- PDSA, *see* PDCA
- Pearson coefficient, 101
- Pie Charts, 50
- Pivot Table, 48
- Point estimators, 73
- Poka-Yoke, 186
- Potential capability, 84
- Power test, 132
- Precision-to-tolerance ratio, 150
- Probability distributions
 - Bernoulli, 67

- Binomial, 67
- Chi-square, 69
- Exponential, 67
- F-distribution, 69
- Geometric, 67
- Normal, 67
- Poisson, 67
- t-distribution, 69
- Probability plots, 70
- Process capability ratio, *see* PCR62
- Process cycle efficiency, *see* PCE18
- Process improvement, 16
- Process sampling, 63
- Process variability, 75
- Project charter, 18, 40
- Q-sort, 119
- Quality, 12
 - Audits, 27
 - Costs, 21
 - Dimensions, 12
 - Health care, 12
 - Health Care Informatics, 14
 - Improvement, 16
 - Service, 12
- Queueing theory, 185
 - Optimal staffing, 185
 - Pooling, 185
 - Triaging, 185
- R&R, 149, 150
- Radar charts, 57
- Random sampling, 63
- Rational subgroups, 64, 194
- Regression, 102, 172
- Repeatability and Reproducibility Analysis, *see* R&R
- Repeated measures, 135
- Report cards, 159
- Residual plots, 104
- Risk analysis, 114
- Risk factors, 17
- Risk priority number, *see also* RPN
- Root cause analysis, 114
- RPN, 121
- Run charts, 209
- Sample statistics
 - Mean, 64
 - Variance, 64
 - Standard deviation, 64
- Scatter diagrams, 19, 97
- Scatter plots, *see* Scatter diagrams19, 72
- Semmelweis, 23
- Sensitizing rules, 196
- Seven deadly diseases of management, 20
- Shewhart, 22, 193
- Shewhart cycle, *see* PDCA
- Sigma level, 28, 31
- SIPOC diagram, 18, 41
- Six-Sigma, 19, 31
- SMART, 41
- Spaghetti diagram, 36
- Spearman coefficient, 101
- Specification limits, 84
- Statistical process control (SPC), 19
- Stem-and-leaf diagrams, 19, 79
- Sum of squares, 64
- Swim-lane diagram, 35
- Swiss-cheese model, 170, 186
- t-test, 131
- Time traps, 46
- Total Quality Improvement (TQM), 19
- Transportation, 183, 184
- Type I error, 132
- Type II error , 132
- Unimodal, 77
- Value-added, 36
- Value-stream map, 36
- Variability, 16
- VOC, 18, 27
- Voice of the Customer, *see* VOC
- Why-Why diagrams, 118
- WIP, 18, 46, 183, 185
- Work sampling, 46
- Work-in-Process, *see* WIP
- Workflow diagrams, 35
- Yield, 31

Z-score, 68
Z-test, 130
zig-zag, 210

