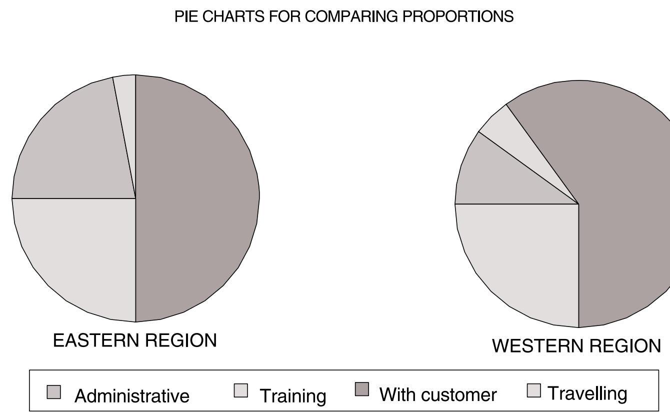
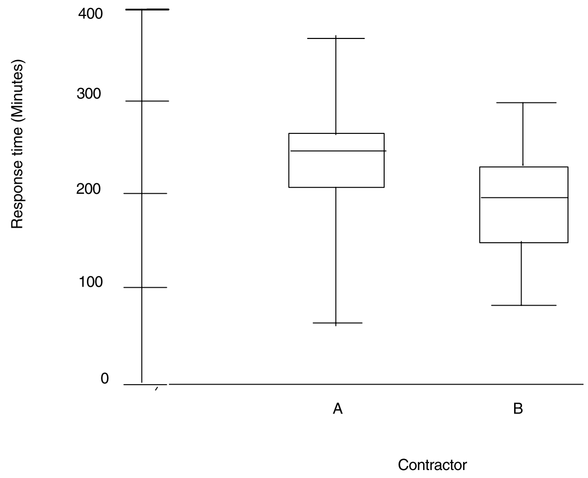
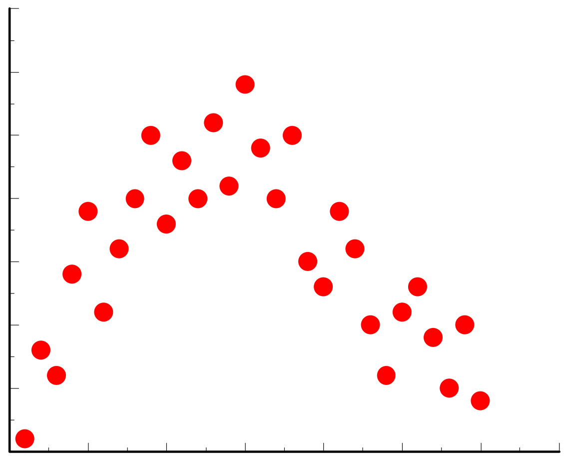
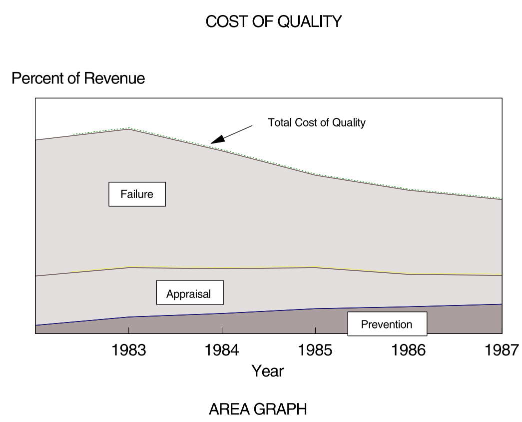




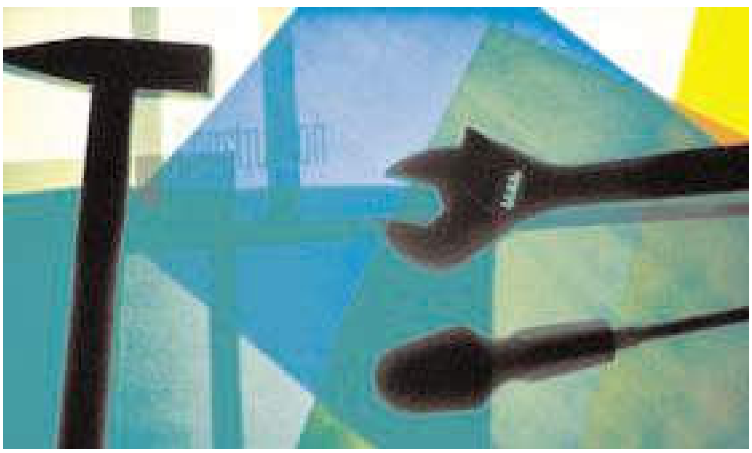
TOOLS FOR QUALITY IMPROVEMENT





The School of Quality

TOOLS FOR QUALITY IMPROVEMENT



ST UNIVERSITY

TOOLS FOR QUALITY IMPROVEMENT

Teams of people working together on projects in order to achieve pre-defined quality objectives or eliminate the root cause(s) of a specific problem are an integral part of ST culture. Our emphasis is to ensure that all these team members are sufficiently equipped to undertake their projects. This requires basic understanding of a methodology which provides a framework to guide them through their project and working knowledge of the more common data collection, analysis and interpretation tools.

This self-training guide is a collection of such tools. You will find that regardless of the methodology or process your team uses to carry out a project, the tools discussed here will allow you to analyze and interpret data in almost all cases.

In discussing each tool, we have adopted the following format:

- *Concepts* introduces the tool and explains general principles involved
- *Example* shows how the tool is used in a problem-solving or quality-improvement context
- *Construction* or *How to* illustrates how to use the tool to visualize a set of data
- *Interpretation* helps you understand the data as illustrated by the tool
- *Avoiding pitfalls* offers some tips on what to avoid when using the tool
- *Questions* allows you to test your knowledge of each tool and the suggested solutions provide you practical tips which may not have been covered in the text.

The material presented in this self-training guide is by no means comprehensive. It is intended to enhance and complement your current knowledge of the tools so that you and your team can use them to effectively tackle your projects.

If you have any comments, suggestions or feedback concerning this self-training guide, please feel free to contact Murali at meempat.murali@st.com

School of Quality, ST University

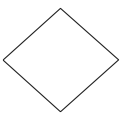
INDEX

1. FLOW DIAGRAM

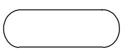
CONCEPTS

A flow diagram is a graphic representation of the sequence of steps that we perform to produce an output. The output may be a physical product, a service, information, or a combination of the three. A flow diagram is used at the beginning of a project to determine the project scope and boundaries; during the project to provide a common understanding of the process under investigation; and towards the project end, when it could be used to make comparisons between current (old) and improved (new) processes. The common symbols used in a flow diagram are:

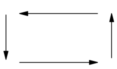
1. The **activity** symbol is a simple rectangle box, which gives a brief description (1-2 words) of an an activity in the process.



2. The **decision** symbol is a diamond shaped box used to denote any point in the process where a decision, such as yes/no, up/down or on/off, must be made, leading to different process branches.

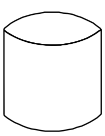


3. The **terminal** symbol is a rather flat rectangle with rounded corners, used to denote the beginning and end of a process or sub-process, often using the words *start*, *begin*, or *end*.

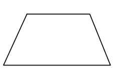


4. **Flow lines** have an arrow at the end, indicating the process flow to the next activity or step. Besides the usual downward flow, flow lines can indicate a process loop (going back to an activity) or a process branch (going to a different set of activities or a sub-process), based on the decision made.

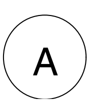
5. The **document** symbol has the shape of a printout or paper and is used when the activity described is a printed form.



6. A **database** symbol is a cylinder and is used when the activity involves access to or retrieval from a database.



7. The **wait** symbol is a trapezoid, providing a means to show where normal processing is halted or suspended to be continued later on.



8. The **connector** or continuity symbol is a circle with a letter or number in it, used to denote a break point in the process (most frequently used when the flow diagram reaches the edge of the paper). An identical symbol is then used to show where the diagram continues.

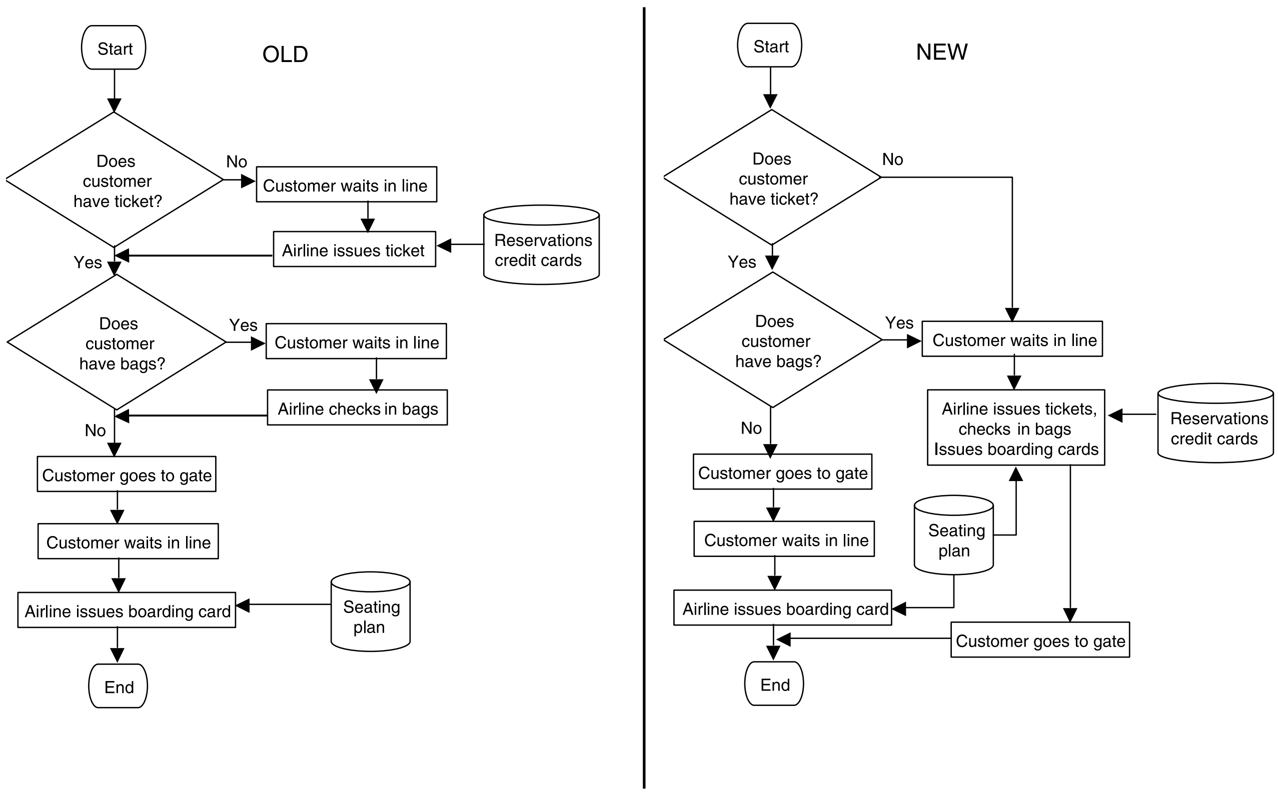
1. FLOW DIAGRAM

EXAMPLE

The example below illustrates how flow diagrams can prove useful in both in the diagnostic and remedial phases of a project. The left hand side of Figure 1 below shows the process that airlines used a few years ago to issue tickets and boarding passes to passengers without ticket. An examination of the diagram shows that passengers must stand in two lines, first to purchase the ticket and then at the gate to board. Giving the ticketing agents access to the seating database is an obvious process improvement that is sure to make passengers happy by eliminating the need to stand in one of the lines. This improvement is shown on the right side of the figure.

Figure 1. Flow diagram of airline process for issuing tickets and boarding passes

Customer arrives at airport needing to purchase ticket or check bag



1. FLOW DIAGRAM

CONSTRUCTION

1. Define how the flow diagram will be used in your project.

2. Decide on desired outcome of the flow diagram: high level, matrix or detailed? Construct a high level diagram before constructing a detailed diagram.

   A high level flow diagram groups all related activities and decisions in blocks or sub-processes, so that the final diagram is a collection of sub-processes linked together from start to finish.

   A matrix (or cross-functional) flow diagram shows the process flow both sequentially and by function. The diagram is divided into rows or columns, each representing a department, function or other sub-unit of the system under study. Each symbol of the flow diagram is placed in the row or column of the unit responsible for executing it.

3. Define boundaries to the process by identifying first and last steps.

4. Document each step in the sequence, starting from the first (or last) step. Draw the process accurately and consistently from the top left to the bottom right edge of the page.

5. When a decision symbol is encountered, select one branch and continue the flow diagram of that branch.

6. When you run into an unfamiliar activity of a sub-process, make a note and continue the flow diagram.

7. Repeat steps 4, 5 and 6 until the last (or first) activity is reached.

8. Go back and flow diagram the other branches of the decision symbols.

9. Review the diagram for any missed decision points or special cases that might cause some work to follow a different process.

10. Discuss how to fill in the unfamiliar steps or areas of the flow diagram, and verify its accuracy (observe the process directly, interview experts, show the diagram to line workers, and so on).

11. Interpret the flow diagram, once it is accurate and complete.

1. FLOW DIAGRAM

INTERPRETATION

The best way to gain an understanding of a process is to walk it through the flow diagram step by step, following the flow indicated by the arrows. To do this:

1. Examine each decision symbol.
- Is this a checking activity?
- Is this a complete check, or do some type of errors go undetected?
- Is this a redundant check?

2. Examine each rework loop.
- Would we need to perform these activities if we had no failures?
- How long is this rework loop (steps, time lost, resources consumed, ...)?
- Does this rework loop prevent the problem from re-occurring?

3. Examine each activity symbol.
- Is this a redundant activity?
- What is the value of this activity relative to its cost?
- How have we prevented errors in this activity?

4. Examine each document or data base symbol.
- Is this necessary?
- How is this kept up to date?
- Is there a single source for this information?

AVOIDING PITFALLS

- When analazying a problem failure, be sure to document the *actual* process, and not how the process *should be*.

- Don't forget to update the flow diagram as new information is uncovered or as the process is modified.

1. FLOW DIAGRAM

QUESTIONS

1. When should you use a flow diagram?

2. What is the symbol for an activity? a decision? a connector?

3. When is a high-level flow diagram best used?

4. How do you proceed when you encounter a branch point when drawing the flow diagram of your process?

5. How can a flow diagram help you better understand a process?

6. What are the common pitfalls to avoid when using a flow diagram as a tool?

See recommended solutions on page 55.

2. BRAINSTORMING

CONCEPT

The difference between creative and logical thinking is not a difference in the truth or usefulness of the ideas produced but rather in the method used to produce the idea. Logical thought follows rules and can be reproduced using the same rules, whereas creative thought is not determined by rules and normally cannot be duplicated by others. Brainstorming is a tool for generating new ideas, using a few simple rules for discussion that increase the chances for creativity and innovation.

HOW TO BRAINSTORM

1. Phrase the statement properly. Focus on the issue but be broad enough for creativity.

2. Prepare.
- Communicate the subject to participants ahead of time
- Determine the ideal number of participants
- Provide an appropriate medium to list all contributions

3. Follow the conceptual rules.
- No criticism or evaluation of any kind
- Be unconventional
- Aim for quantity of ideas
- "Hitch-hike" on other ideas

4. Follow the practical rules.
- Make contributions in turn with only one idea per turn. You may pass if necessary.
- Do not provide explanation of ideas.

5. Facilitate the brainstorming session.
- Explain the issue and write it down where everyone can see it
- Have a scribe write all contributions where they can be seen by everyone
- Stop before fatigue sets in

6. Process the ideas.
- Clarify each contribution
- Combine and group similar ideas
- Agree on evaluation criteria

If the subject is project selection, collect data on the severity of problem and contributions to it. If the subject is theories, go to the cause-effect diagram. If the subject is remedies, evaluate ideas in context of time, costs, probability of success, and so on. If the subject is resistance or culture, develop strategies for the serious ones.

2. BRAINSTORMING

INTERPRETATION

Brainstorming is a proven way of generating many ideas on a topic and increasing the creativity of the participants. The resulting lists of ideas usually contain more new and innovative ideas than lists obtained in other ways.

The next step in analyzing the ideas of a brainstorming is to narrow down the list to the vital few, which the team could comfortably focus on. This narrowing or prioritizing of ideas can take the form of voting, a Pareto analysis or consensus.

AVOIDING PITFALLS

- While brainstorming is a great way to generate lists of ideas, it should not be used as a substitute for data.

- The most common barriers to the flow of ideas and creativity are: the inability of the group to suspend judgment and analysis until the list of ideas has been completed and the dominance of one or a few individuals in presenting ideas.

2. BRAINSTORMING

QUESTIONS

1. What is the first step in a proper brainstorming session?

2. Name at least three conceptual and three practical rules to observe during a brainstorming session.

3. How are the large number of ideas generated in a successful brainstorming session handled?

4. Is there a need to collect data to support ideas and theories generated from brainstorming session? Why/why not?

5. How can a team avoid the dominance of one or a few individuals during a session?

See recommended solutions on page 56.

3. CAUSE-EFFECT DIAGRAM

CONCEPTS

A cause-effect diagram is an effective way to organize and display theories about what the root causes to an observed symptom might be. It illustrates the causal relationships involved between the various identified contributions and the observed symptom or effect.

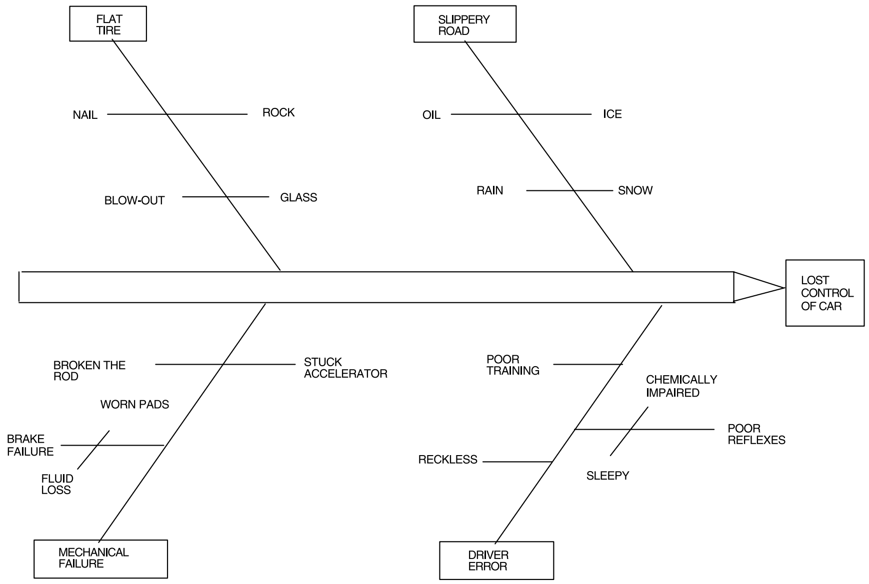
This tool has three prominent basic features:

- It is a visual representation of the factors that might contribute to an observed effect or phenomenon that is being examined.
- The interrelationships among the possible factors are clearly shown. One causal factor may appear in several places in the diagram. For example, if temperature affects both moisture content and physical dimensions, then temperature would appear in both places.
- The interrelationships are generally qualitative and hypothetical. A cause-effect diagram is usually prepared as a prelude to developing the data needed to establish causation empirically.

The most important consideration in the construction of a cause-effect diagram is the clear understanding of the cause-effect relationship. All possible sources of causes need to be considered. In addition to the four W's (What, When, Where, Who), there are two other lists that can help you remember to consider several classes of possible causes for a problem. These lists are characterized as the five M's (Man, Machine, Method, Material, Measurement) in manufacturing and the five P's (People, Process, Place, Provision, Patron) in services.

EXAMPLE

Figure 2.Cause-effect diagram for lost control of car



3. CAUSE-EFFECT DIAGRAM

In Figure 2, the phenomena to be explained is "lost control of car." Some of the possible factors contributing to lost control are a flat tire, a slippery road, mechanical failure, and driver error. Each of these major categories of causes may, in turn, have multiple causes. A flat tire may come from a nail, a rock, glass, or a blow-out from material failure. The causal relationship can be traced back still more steps in the causal chain if necessary or appropriate. For example, loss of control may arise from a mechanical failure. This failure may be a brake failure, which, in turn, may come either from fluid loss or from worn brake pads, and so on.

CONSTRUCTION

1. Define the effect or symptom for which the causes must be identified.

2. Place the effect or symptom being explained at the right, enclosed in a box. Draw the central spine as a thick line pointing to it.

3. Use brainstorming or a rational, step-by-step approach to identify possible causes.

4. Each of the major cause areas should be placed in a box and connected with the central spine by a sloping line (thereby forming a branch).

5. Add causes for each main area.

6. Add subsidiary causes for each cause already entered.

7. Continue adding possible causes until each branch reaches a root cause.

8. Check the logical validity of each causal chain.

9. Check for completeness.

INTERPRETATION

It is important to distinguish between theory and fact. Cause-effect (C-E) diagrams present and organize theories. Only when theories are tested with data can we prove causes of observed phenomena.

When interpreting a cause-effect diagram, your primary task is to check for completeness. To do so, consider the following points:

- Be certain you have at least asked how each of the four W's and each the five M's or five P's might apply to the effect.
- Generally, each main branch of the diagram will have at least three or four additional branches.
- If some main branches have substantially fewer causes effected to them, or if the causes on them do not go back as many steps in the causal chain, you may not have as full an understanding of that element of the process.

- Verify that the cause at the end of each causal chain is potentially a root cause by checking if you can trace a logical causal relationship from that cause, through all its intermediate causes, to the final effect being explained. That cause is, in principle, directly controllable. Therefore, if shown to be true, that cause could be eliminated, and the effect would disappear or be reduced.

AVOIDING PITFALLS

- Be careful not to confuse the orderly arrangement of theories found in a cause-effect (C-E) diagram with the real data obtained through empirical testing.
- Be sure to test each causal relation in the cause-effect diagram for logical consistency to ensure its usefulness.
- Construct the diagram only once the symptoms have been analyzed as thoroughly as existing information will permit, which will help avoid a large, complex diagram that can be difficult to use.
- Do not limit the theories that are proposed and considered or you may unintentionally mask or ignore the ultimate root cause(s).

3. CAUSE-EFFECT DIAGRAM

QUESTIONS

1. What is the primary purpose of a cause-effect diagram?

2. In a typical cause-effect diagram, what description is placed in the box on the extreme right?

3. Which common tool could a team use to generate inputs for their cause-effect diagram?

4. Name the two checking activities to do immediately after completing the diagram.

5. What needs to be done after analyzing a cause-effect diagram?

6. What is the danger of limiting the input of ideas when drawing the diagram?

See recommended solutions on page 56.

4. DATA COLLECTION

CONCEPTS

Data and facts are fundamental to every quality improvement effort. Without the facts, our problem solving efforts are reduced to a guessing game with relatively low odds of success. Therefore, it is fundamental to gather good data using simple data collection tools.

Information generation begins and ends with questions. To generate information, we need to:

- formulate precisely the question we are trying to answer
- collect the data and facts relating to that question
- analyze the data to determine the factual answer to the question
- present the data in a way that clearly communicates the answer to the question.

In order to generate useful information, planning for good data collection proceeds along the following lines:

- What question do we need to answer?
- How will we recognize and communicate the answers to the question?
- What data analysis tools (Pareto diagram, histogram, bar, graph, etc...) do we envision using, and how will we communicate the results?
- What type of data do we need in order to construct this tool and answer the question?
- Where in the process can we get this data?
- Who in the process can give us this data?
- How can we collect this data from these people with minimum effort and chance of error?
- What additional information do we need to capture for future analysis, and traceability?

HOW TO COLLECT DATA

1. Formulate questions that relate to specific information needs of the project.

2. Consider appropriate data-analysis tools and ensure the data needed for the analysis are being collected.

3. Define comprehensive data collection points.

4. Select unbiased collectors; understand the environment in which the data is being collected.

5. Design data collection forms using the KISS motto: Keep It Short and Simple.
   - Reduce opportunities for errors.
   - Capture data for analysis, reference and traceability.
   - The form should be self explanatory.
   - Incorporate a professional look.

6. Prepare the instructions for collecting data

7. Test the form and instructions and train the data collectors.

8. Audit the collection process and validate the results.

INTERPRETATION

Data should be used to answer the questions for which it was collected using the tools identified while planning for the collection. The following are types of biases that may occur and which you should be aware of. Some can be avoided with careful planning, while others require attention during analysis.

- Exclusion bias: Some part of the scope of the process being investigated has been left out.
- Interaction bias: The process of collecting the data itself may affect the process being studied.
- Perception bias: The attitudes and beliefs of the data collectors can sometimes influence what they see and how they record it.
- Operational bias: Failure to follow the established procedures is the most common operational bias.
- Non-response bias: Missing data can bias the result. The fact that they are missing is a clue that they are different from the rest in some way.
- Estimation bias: The formula and methods used to calculate statistics from the collected data may give certain types of biases.

AVOIDING PITFALLS

Avoid using data collection when creativity is important: brainstorming theories, possible remedies, or cultural issues. Before collecting new data, ensure that the answer cannot be found already using existing data.

4. DATA COLLECTION

QUESTIONS

1. Why does data collection constitute an integral part of the problem solving process?

2. What are five key considerations when planning to collect data for your project?

3. What are the major considerations when designing a data collection form?

4. A common bias to avoid in data collection is the non-response bias. What is this?

5. Discuss the merits and disadvantages of using existing data as opposed to collecting new data.

See recommended solutions on page 57.

5. GRAPHS AND CHARTS

CONCEPTS

Graphs and charts are pictorial representations of quantitative data. They can summarize large amounts of information in a small area and communicate complex situations concisely and clearly.

On line and bar graphs, the horizontal axis portrays the independent variable. The vertical axis portrays the dependent variable. The independent is the one that is believed to cause or determine the dependent one. On a single line graph, there is only one point for each independent variable value. There can be multiple points with the same dependent variable value.

Each type of graph has its own particular strength.

- Line graphs are good for showing trends.
- Bar graphs are particularly useful for showing comparisons among categories.
- Pie charts are useful principally for showing proportions.

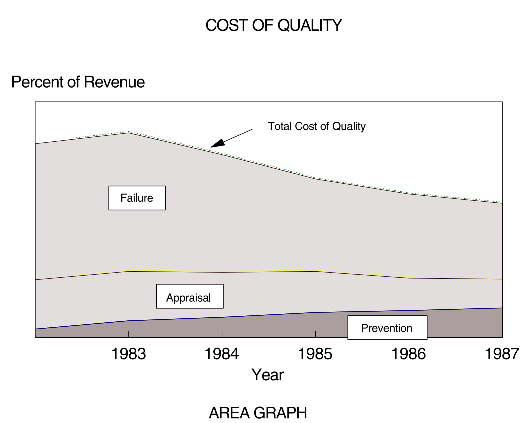
EXAMPLES

Figure 3. Trend line graph

Figure 3 illustrates a very clean, easy-to-read line graph (also known as a trend line graph). The vertical axis shows the number of non-conforming components in parts-per-million over a period of months.

Figure 4. Area graph

An area graph is a line graph which presents several lines on the same picture, so that you can read both the values for each of the lines and the cumulative value at any given point of the vertical axis. The area graph in Figure 4 represents the total costs of quality. The first line shows the costs of prevention; the second line shows costs of prevention plus appraisal; the third shows costs of prevention plus appraisal plus failure. Correspondingly, the area between the second and first lines, and the third and second lines, represent the costs of appraisal and costs of failure respectively. What conclusions can you draw from this figure?



5. GRAPHS AND CHARTS

Figure 5.Bar graph
Bar graphs can be drawn grouped or stacked. Figure 5 shows grouped bars where the height of each bar represents the number of leaks in roller assemblies that occurred in tractors produced in a particular year. For each year, there are three differently shaded bars, representing three models of equipment. This chart allows you to make two types of comparisons: the number of leaks for each model and the relative number of leaks among models.

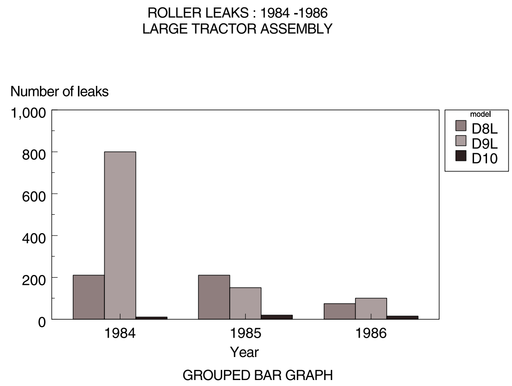
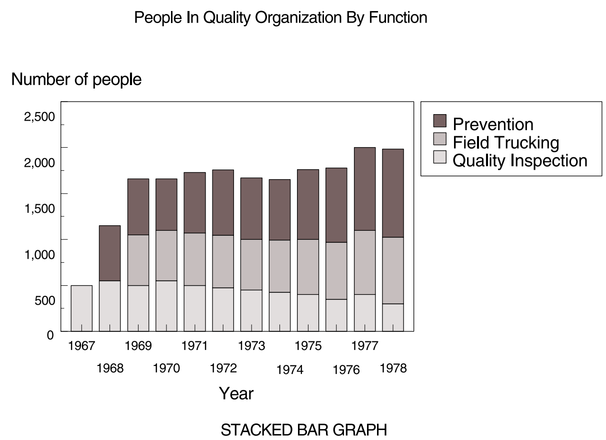


Figure 6. Stacked bar
Figure 6 shows a stacked bar, representing the number of people engaged in various quality activities over different years. Each bar shade represents a specific type of quality activity.



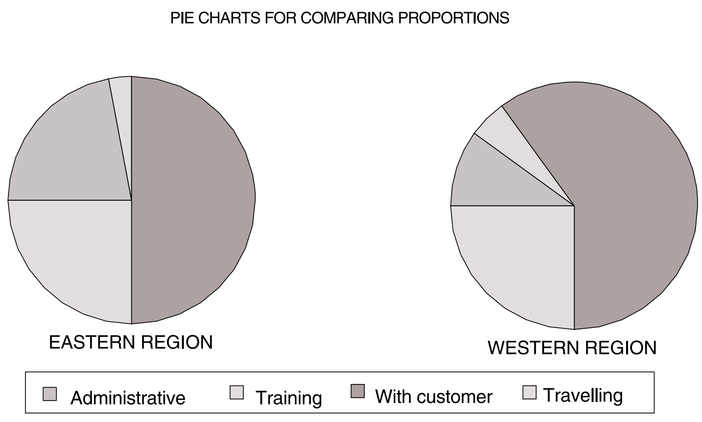


Figure 7. Pie chart
Our last example shows a pie chart of how the sales staff of two different organizations spend their time. The area of each task is proportional to the time spent on that task. The differences between the two regions stand out clearly. The Western region sales staff spends substantially less time on administrative tasks and thereby, can spend more time with customers and in training.

5. GRAPHS AND CHARTS

CONSTRUCTION

Line graphs

1. Determine the range of the vertical axis and the size of each increment. Label the axis, including the unit of measure.
2. Determine the range of the horizontal axis and the size of each increment. Label the axis.
3. Draw the axes and plot in each data point; connect the points with a line.
4. Label and title the graph.

Bar graphs

1. Determine the range of the vertical axis and the size of each increment. Label the axis, including the unit of measure.
2. Select an appropriate bar graph - simple, grouped or stacked.
3. Determine the number of bars. Draw and label the horizontal axis.
4. Determine the order of the bars; draw them on the graph.
5. Label and title the graph.

Pie charts

1. Determine the percentage of each category
2. Convert the percentage values into degrees of angle.
3. Draw a circle and carve out the segments of the pie chart.
4. Label the segments and title the chart.

INTERPRETATION

Keep in mind these guidelines when constructing graphs:

- Graphic integrity: Construct your graph so that it represents your actual data. Use data labels to reinforce graphic information and ensure that the true information is conveyed.

- Consistent scale: Numeric scales must maintain regular intervals. Different graphs that might be compared to each other should all be drawn to the same scale.

- Ease of reading: If a graph is easy to read, it will be better understood and remembered. Labels, placed closely to the object they identify, improve clarity.

- Consistency of symbols: When two or more graphs are to be compared, use the same symbols to minimize confusion in interpreting the graphs.

- Simplicity: Be sure not to obscure information with unnecessary lines, colors, 3-D, shading and other decoration. Text and decoration should add value or information.

5. GRAPHS AND CHARTS

AVOIDING PITFALLS

- Be sure your graph does not distort reality by choosing the appropriate scale.
- Use consistent scales from one chart to the next when they are to be compared.
- Be careful not to vary the scale on one axis.
- Keep charts and graphs simple, avoiding excessive decoration.

5. GRAPHS AND CHARTS

QUESTIONS

1. When is a line graph best used? A bar graph? A pie chart?

2. In a line graph, what are the two principal axes called? Which one is placed horizontally and which one is placed vertically?

3. What is the difference between a grouped bar and a stacked bar presentation?

4. What are the key concepts to bear in mind when preparing a graphic chart?

5. How does varying the scale on one axis affect the presentation of a line graph?

See recommended solutions on page 58.

6. PARETO ANALYSIS

CONCEPTS

Pareto analysis is a ranked comparison of factors related to a quality problem. It helps quality improvement project teams identify and focus on select vital factors.

In the early 1950s, Dr. Joseph Juran noted a universal phenomenon that has come to be called the Pareto principle. The Pareto principle states that in any group of actors contributing to a common effect, only a relative few account for the bulk of the effect. Juran has also coined the terms "vital few" to refer to the small number of contributions which account for the bulk of the effect, and "useful many" to refer to all the other factors which account for a smaller proportion of the effect.

Pareto diagrams and tables are presentation techniques used to show the facts and separate the vital few from the useful many. They are widely used to help quality improvement teams and quality councils make key decisions at various points in the quality improvement or problem-solving sequence.

One key application of Pareto charts is to make comparisons between "before" and "after" charts to show the effect or impact of remedial or improvement actions in the "vital few" areas identified.

Well-constructed Pareto diagrams and tables include three basic elements:
- Contributors to the total effect, ranked by the magnitude of their contribution.
- Magnitude of the contribution of each factor expressed numerically.
- Cumulative-percent-of-total effect of the ranked contributors.

EXAMPLE

A quality improvement team was formed to improve the format of forms coming from field sales offices to the home office. There were 18 items on the order form, designated A-R. The team developed a check sheet which it used to collect the frequency of errors on the forms for one week. The results of the study, in the form of a Pareto table is shown in Figure 8.

The team then went on to plot the Pareto diagram shown in Figure 9. The implications of the analysis is profound. If the team can resolve the causes creating errors for the four vital few information items, they can significantly improve the quality of the order forms coming from sales offices.

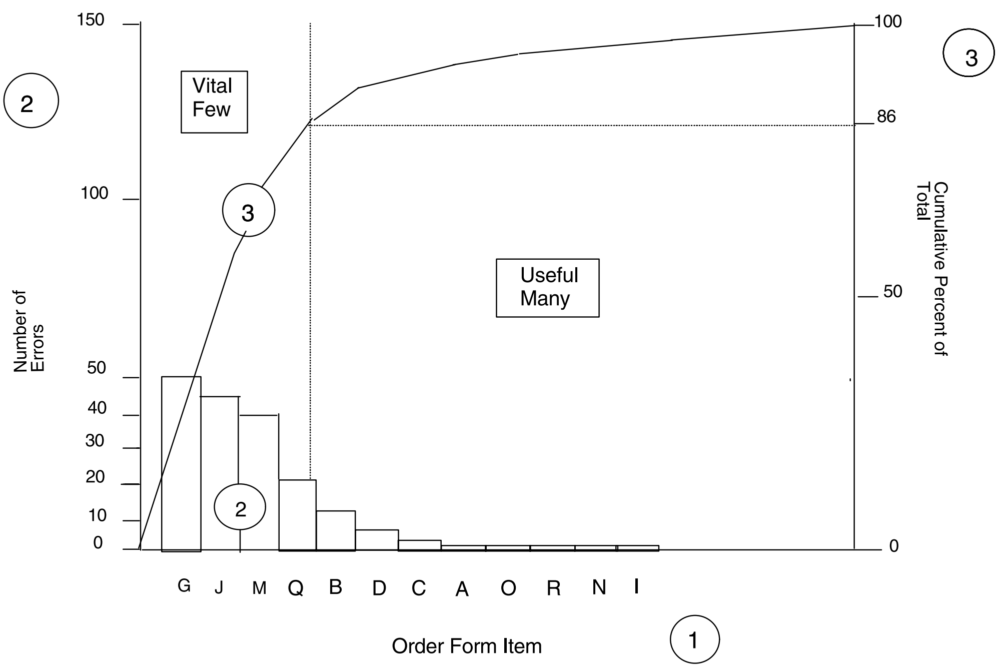
This is an important point: without the facts and without a Pareto analysis, the team would have been faced with the much larger and more costly task of trying to find ways to prevent errors from occurring on all 18 items. Thus, we can clearly see from a Pareto table or chart that significant improvement can be achieved with a smaller, but more precisely focused effort.

6. PARETO ANALYSIS

Figure 8. Pareto table of errors on order forms

① Order-Form Item	② Number of Errors	Percent of Total	③ Cumulative-Percent of Total
G	44	29	29
J	38	25	54
M	31	21	75
Q	16	11	86
B	8	5	91
D	5	3	95
C	3	2	97
A	1	0.67	98
O	1	0.67	98
R	1	0.67	99
N	1	0.67	99
L	1	0.66	100
I	0	0	100
E	0	0	100
H	0	0	100
K	0	0	100
F	0	0	100
P	0	0	100
TOTAL	150	100%	

Figure 9. Pareto chart



CONSTRUCTION

1. Total the data on the effect of each contributor, and sum these to determine the grand total.

2. Re-order the contributors from the largest to the smallest.

3. Determine the cumulative percent of total.

4. Draw and label the left vertical axis (label the axis from 0 to the grand total or just beyond).

5. Draw and label the horizontal axis:
   - Divide the axis into as many divisions as there are contributors.
   - List the contributors from the largest to the smallest going from left to right.
   - Label each contributor.
   - If captions are too long, label the axis A, B, C ... and provide a separate key.

6. Draw and label the right vertical axis (label the axis from 0 to 100%). Line up the 100% mark with the grand total on the left axis and label the axis (for example, "Cumulative Percent of Total").

7. Draw bars to represent the magnitude of each contributor's effect. The height of each bar corresponds to the magnitude of the contribution as measured on the left axis.

8. Draw a line graph to represent the cumulative percent of total. The plotted points correspond to the cumulative percent as measured on the right axis. As a check, the cumulative percent of total for the first ranked contributor should be even with the height of the first bar.

9. Title the chart.

6. PARETO ANALYSIS

INTERPRETATION

The goal of Pareto analysis is to separate the numerous problems or causes of problems into two categories: the vital few and the useful many. The easiest way to do this is to look for a break point in the slope of the cumulative percentage of total line graph on the Pareto diagram.

Determining the break point is not an exact science. In practice, a quality improvement team faced with interpreting a Pareto diagram that does not show a clear break point usually takes the following approach:

- Identify those few contributors which account for about 60% of the quality effect.
- Call these the "vital few", and begin the diagnostic journey.
- When the diagnostic and remedial journeys are completed for these vital few, repeat the Pareto analysis. The contributors that were in the awkward zone - not clearly part of the vital few or the useful many - may now be among the vital few.
- Repeat steps 1 through 3 as long as profitable projects can be identified.

AVOIDING PITFALLS

- The most common mistake with Pareto analysis is failure to use the technique. This results in unfocused, low-payoff quality improvement efforts.
- Pareto analysis is most powerful when we use objective data and facts, rather than opinions, to rank the order of the categories.
- Sometimes the data indicate no clear distinction between the categories. In this situation, all the bars on the Pareto diagram are roughly the same height or it takes more than half of the categories to account for more than 60% of the quality effect.

6. PARETO ANALYSIS

QUESTIONS

1. How is a Pareto chart different from a normal bar chart?

2. State the rationale behind the Pareto or 80/20 principle.

3. Name the three elements in a Pareto chart.

4. How can the Pareto tool be used to illustrate improvements in a targeted area?

5. After constructing a Pareto diagram, what steps can a team take if no obvious distinctions between the various categories appear?

See recommended solutions on page 59.

7. SCATTER DIAGRAM

CONCEPTS

A scatter diagram is a graphic representation of the relationship between two variables. In quality improvement, scatter diagrams are usually used to test cause-effect relationships in the diagnostic journey.

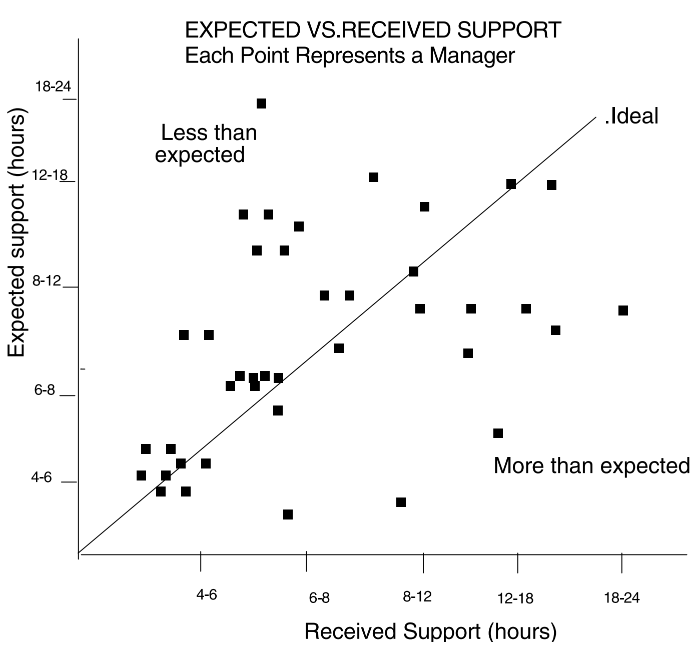
- Concept 1: Discovering true cause-effect relationships is a key to effective problem solving.
- Concept 2: Cause-effect relationships almost always show variations.
- Concept 3: It is easier to see the relationship in a scatter diagram than in a simple table of numbers.

EXAMPLE

Managers in a large corporation were complaining about the lack of administrative support services received from a central pool offering special on-call administrative services. The team chartered to investigate this started with the hypothesis that the managers' expectations were being met but that they were unaware of it since the time was broken up into a few hours one day and a few hours the next.

To investigate the validity of the complaint, the team surveyed the managers and asked them to specify the number of hours a week of support they expected. The team then reviewed time records to see how much support each manager actually received. The scattered diagram of this paired data is shown below:

Figure 10. Scatter diagram



The diagonal line in figure 10 is a target line indicating when "received support" equals "expected support". Ideally, the team hoped to see all points falling close to the line. Points above the line indicated that the managers' expectations were not being met. Points below the line indicated that they were receiving more support than they realized. The fact that more points fell above than below it forced the team to conclude that there was more to the complaints than a simple misconception about the amount of service received.

7. SCATTER DIAGRAM

CONSTRUCTION

1. From the table of raw data, determine the maximum and minimum values for each variable.

2. Decide which variable will be plotted on the horizontal axis (if showing a cause-effect relationship, put the suspected cause variable on the horizontal axis).

3. Draw and label the horizontal and vertical axes. For each label, make the lowest label slightly lower than the minimum value and the highest label slightly higher than the maximum value.

4. Plot the paired data. Use concentric circles to indicate identical paired-data points.

5. Title the chart and provide other appropriate notations.

INTERPRETATION

Analyzing scatter diagrams is a 4-step process.

1. Develop a plausible and relevant theory about the suspected relationship between 2 variables of interest.
2. Collect appropriate paired data and construct a scatter diagram.
3. Identify and classify the pattern of correlation.
4. Question your original theory and consider the other explanations for the observed pattern of correlation.

General interpretation of each type are as follows (see Figures 11a-f):

- Strong, positive correlation: The value of Y clearly increases as the value of X increases.

- Strong negative correlation: The value of Y clearly decreases as the value of X increases.

  Because the points are so conspicuously grouped around an imaginary line (the trend line) through the middle of the mass of the points, we are very confident of the relationship between X and Y. This suggests that the system under study is one where control of one variable results in control of the other.

- Weak, positive correlation: The value of Y increases somewhat as the value of X increases.

- Weak negative correlation: The value of Y decreases somewhat as the value of X increases.

- Complex correlation: The value of Y appears to be related to X, but the relationship is not simple or smooth. The relationship between X and Y requires examination.

- No correlation: For any value of X, Y can have both large and small values. There does not appear to be any particular relationship between X and Y. Look for other variables that influence Y.

7. SCATTER DIAGRAM

Figure 11a. Strong positive correlation

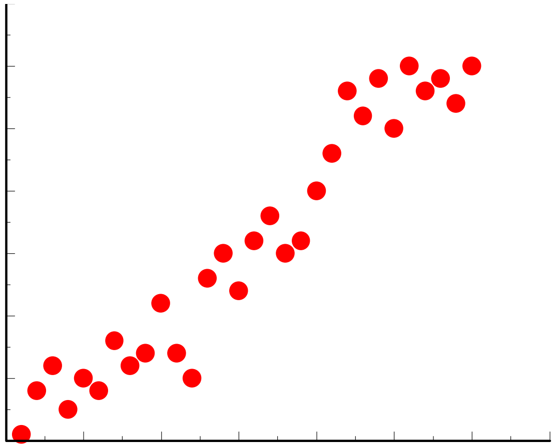


Figure 11b. Strong negative correlation

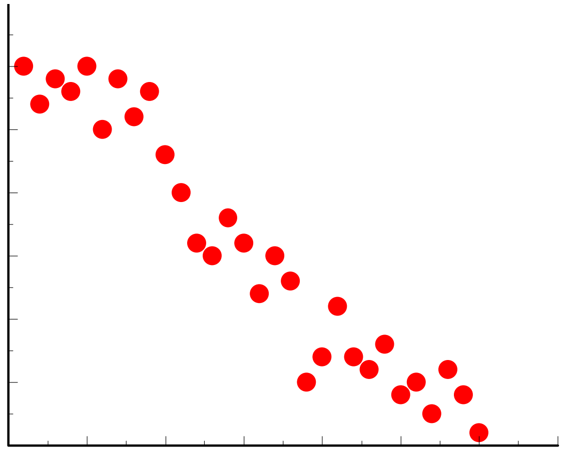


Figure 11c. Complex correlation

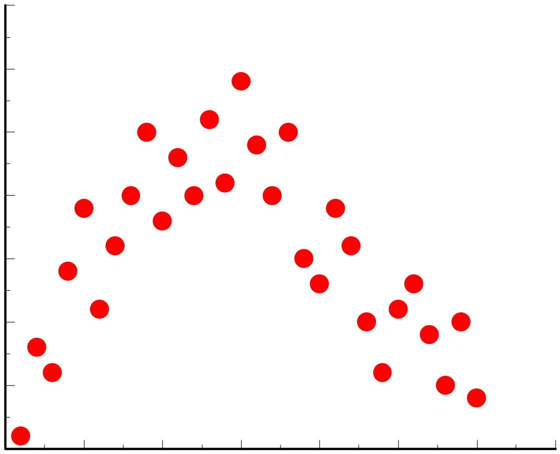


Figure 11d. Weak positive correlation

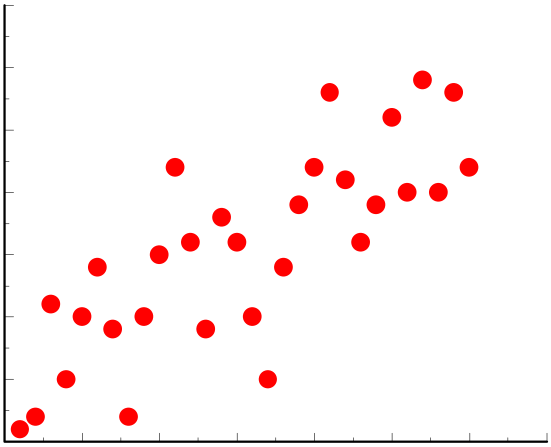


Figure 11e. Weak negative correlation

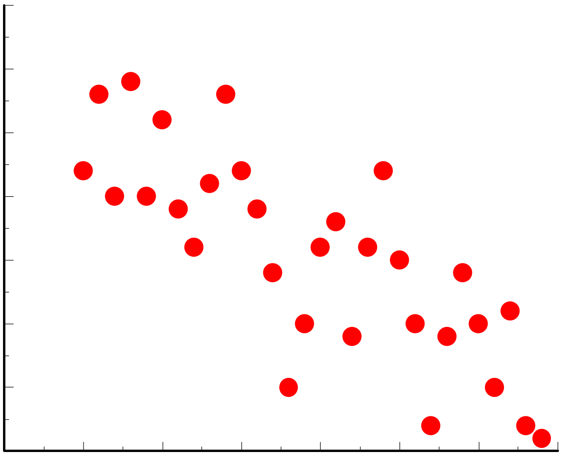
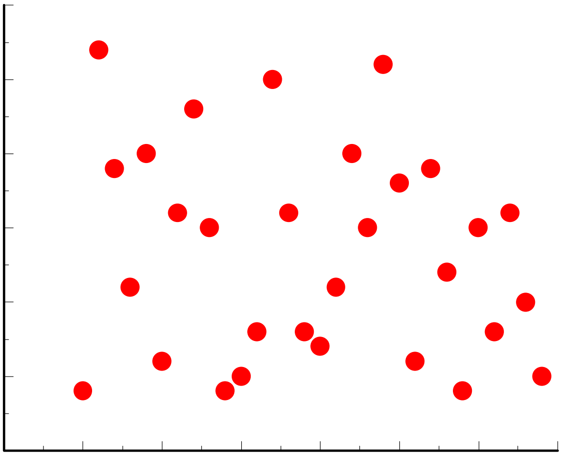


Figure 11f. No correlation



AVOIDING PITFALLS

Range of the data
Restrict interpretation of the scatter diagram to the range of observations. Be careful not to draw conclusions beyond the range of the data.

Range of operation
Be equally wary of the value of gathering and analyzing data beyond the range of the operation of the process.

Effect of scale
Proper interpretation of a scatter diagram depends on careful attention to construction details. Generally, you should draw the axes of scatter diagrams such that (1) the maximum and minimum values on each axis correspond closely to the maximum and minimum values of each variable, and (2) the axes are roughly the same length.

Numerical summary
A number of computer programs print numerical summaries, such as the correlation coefficient and equations of regression lines, without providing a graphic picture of the data. Numerical summaries, by themselves, can be very misleading.

Confounding factors
With a scatter diagram, we study one cause-effect pair among many. It is therefore possible that the correlation we observe in our scatter diagram is due to a cause other than the one we are plotting. This other unseen cause is sometimes called a confounding factor.

No correlation with physical understanding
Scatter diagrams only show relationships; they do not prove cause and effect. You must have a plausible, physical explanation to establish cause and effect.

Absence of stratification
Stratification can make a correlation pattern appear, or it can completely change our interpretation of an apparent correlation.

7. SCATTER DIAGRAM

QUESTIONS

1. What is the primary purpose of a scatter diagram?

2. The variables involved in a scatter diagram are sometimes called the (suspected) cause variable and the (observed) effect variable? Which is to be plotted on the horizontal axis? Which is to be plotted on the vertical axis?

3. Explain the following patterns: a strong positive relationship, a weak negative relationship, no relationship.

4. What is the danger of interpreting a diagram beyond the range of the data from which it was constructed?

5. Scatter diagrams are effective to the extent that they prove or disprove a theory or physical explanation. Where does this theory originate?

See recommended solutions on page 60.

8. HISTOGRAM

CONCEPTS

A histogram is a graphic summary of variation in a set of data. The pictorial nature of the histogram enables us to see patterns that are difficult to see in a simple table of numbers.

Concept 1: Values in a set of data almost always show variation.

Concept 2: Variation displays a pattern.

Concept 3: Patterns of variation are difficult to see in simple tables of numbers.

Concept 4: Patterns of variation are easier to see when the data are summarized pictorially in a histogram.

EXAMPLE

This example illustrates how a histogram could be used to interpret data to verify the existence of a problem. A manufacturer of electronic telecommunications equipment was receiving complaints from the field about low volume sound on long distance connections. Bettina in France couldn't hear her mother in Germany.

A string of amplifiers manufactured by the company was used to boost the signal at various points along the way in these long connections. The boosting ability of the amplifiers (or, the "gain") was therefore a prime suspect in the case.

The design of the product has called for a gain of 10dB. This means that the output from each amplifier should be about 10 times the input signal. This amplification makes up for the natural fading of the signal over long distances. Recognizing that it is impossible to make every amplifier with a gain of exactly 10dB, the design allowed the amplifiers to be accepted if the gain fell between 7.75dB and 12.25dB. Since there are literally hundreds of amplifiers boosting the signal along the way, low gain amplifiers would be compensated by high gain ones to give an acceptable volume.

The QI team investigating the problem arranged to measure the gain of 120 randomly picked amplifiers. The measurements made are shown in Figure 12. The data in this form is not easy to analyze; there are 120 numbers to examine. More importantly, since all the measurements fell within the 7.75-12.25dB range, it is tempting to conclude that the data was of little value.

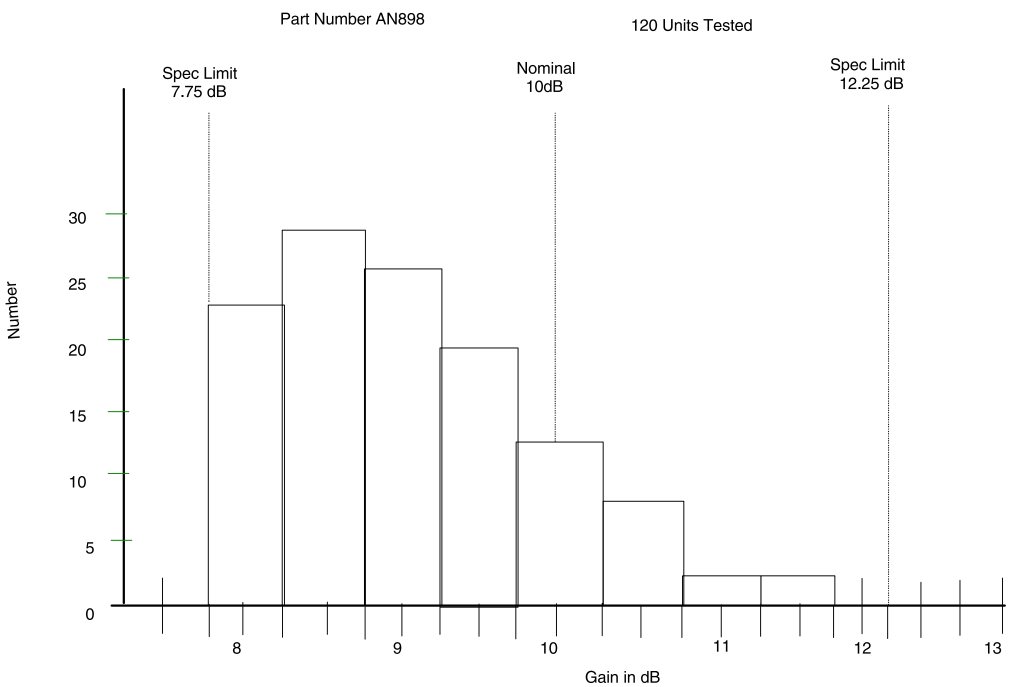
The team decided to construct a histogram to obtain a different perspective (see Figure 13). It becomes immediately obvious that while all the amplifiers fell within the range set, they were certainly not evenly distributed around 10dB. Most of the amplifiers had a lower than nominal value of 10dB. If most of the amplifiers in a long distance connection were boosting their signals a little bit less than expected, the net result would be a low volume level. In other words, Bettina won't be able to hear her mother in Germany.

8. HISTOGRAM

Figure 12. Gain of 120 tested amplifiers

8.1	10.4	8.8	9.7	7.8	9.9	11.7	8	9.3	9
8.2	8.9	10.1	9.4	9.2	7.9	9.5	10.9	7.8	8.3
9.1	8.4	9.6	11.1	7.9	8.5	8.7	7.8	10.5	8.5
11.5	8	7.9	8.3	8.7	10	9.4	9	9.2	10.7
9.3	9.7	8.7	8.2	8.9	8.6	9.5	9.4	8.8	8.3
8.4	9.1	10.1	7.8	8.1	8.8	8	9.2	8.4	7.8
7.9	8.5	9.2	8.7	10.2	7.9	9.8	8.3	9	9.6
9.9	10.6	8.6	9.4	8.8	8.2	10.5	9.7	9.1	8
8.7	9.8	8.5	8.9	9.1	8.4	8.1	9.5	8.7	9.3
8.1	10.1	9.6	8.3	8	9.8	9	8.9	8.1	9.7
8.5	8.2	9	10.2	9.5	8.3	8.9	9.1	10.3	8.4
8.6	9.2	8.5	9.6	9	10.7	8.6	10	8.8	8.6

Figure 13. Histogram with results of amplifier gain testing



8. HISTOGRAM

CONSTRUCTION

1. From the table of raw data, determine the range, high and low values.
   (Range = high value - low value).

2. Decide on the number of cells to use from the following table:

# Data Points	20-50	51-100	101-200	201-500	501-1000	>1000
# Cells to use	6	7	8	9	10	11-20

   Note: Consider using a box plot if the total number of data points is less than 40, unless this is a result of data stratification.

3. Calculate the approximate cell width (cell width = range / # cells)

4. Round the cell width to a convenient number: make the cell width 1, 2, or 5; 0.1, 0.2 or 0.5; 10, 20, 50 etc.

5. Construct the cells by listing the cell boundaries. The first cell should include the lowest value. Cell boundaries should be one more significant digit than the data.

6. Tally the number of data points in each cell. Check that the total tally equals the number of data points.

7. Draw and label the horizontal axis. The range on this axis should include one cell width beyond the lowest and highest value.

8. Draw and label the vertical axis. Label it from 0 to a multiple of 5 above the highest value.

9. Draw in the bars to represent the number of data points in each cell. The height of the bars should equal the number of data points in that cell as measured on the vertical axis.

10. Title the chart (include total number of data points). Show nominal values and limits if applicable.

INTERPRETATION

We know that the values in any set of data will vary and that variation will display some pattern. The goal of our analysis of a histogram is to:

- identify and classify the pattern of variation
- develop a plausible and relevant explanation for the pattern.

The explanation must be based on the team's knowledge and observation of the specific situation, and it must be confirmed through additional analysis.

Figures 14a-h show common patterns of variations:

- The Bell-Shaped Distribution is the normal, natural distribution of data from a normally operating process.
- The Double-Peaked Distribution is usually a combination of two bell-shaped distributions and suggests that two distinct processes are at work.
- The Plateau Distribution is likely to be the result of many different bell-shaped distributions with centers spread evenly throughout the range of the data.
- The Comb Distribution typically indicates data errors.
- The Skewed Distribution typically occurs when a practical limit, or a specification limit, exists on one side and is relatively close to the nominal value. Such skewed distributions are not inherently bad, but a team should question the impact of the values in the long tail.
- The Truncated Distribution is often a smooth, bell-shaped distribution with a part of the distribution removed, or truncated, by some external force, such as screening, 100% inspection, or a review process. These truncation efforts are an added cost and are, therefore, good candidates for removal.
- The Isolated-Peak Distribution suggests that two distinct processes are at work. But the small size of the second peak indicates an abnormality that doesn't happen often or regularly.
- The Edge-Peaked Distribution occurs when the extended tails of the smooth distribution has been cut off and lumped into a single category at the edge of the range of the data. This shape very frequently indicates inaccurate recording of the data.

AVOIDING PITFALLS

- Be careful not to use histograms in an inappropriate context.
- On the other hand, be sure to recognize and use this tool in appropriate situations.

8. HISTOGRAM

Figure 14a. Bell shaped histogram

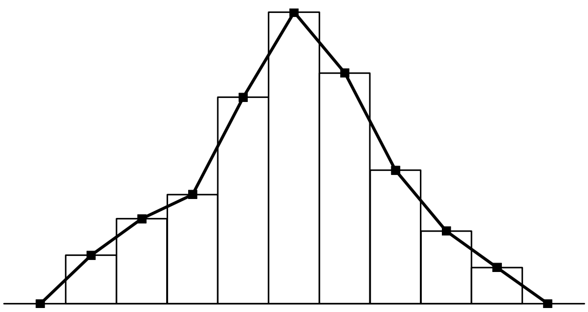


Figure 14b. Double-peaked histogram

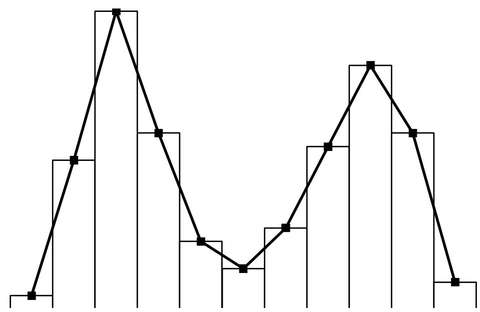


Figure 14c. Plateau histogram

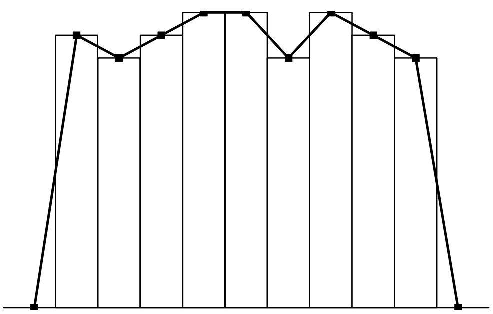


Figure 14d. Comb histogram

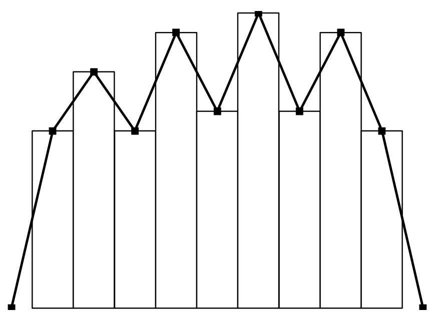


Figure 14e. Skewed histogram

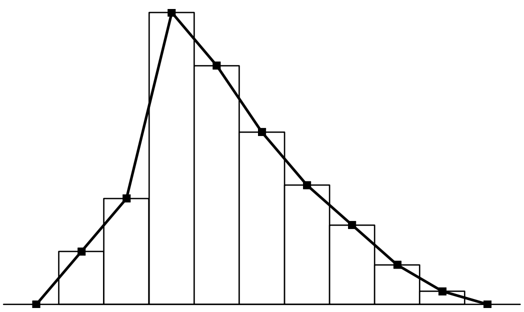


Figure 14f. Truncated histogram

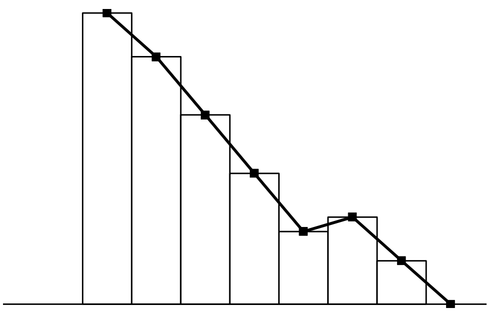


Figure 14g. Isolated peaked histogram

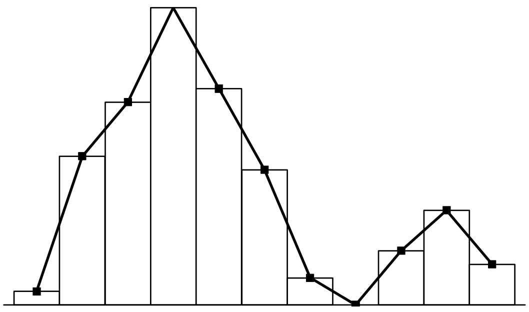
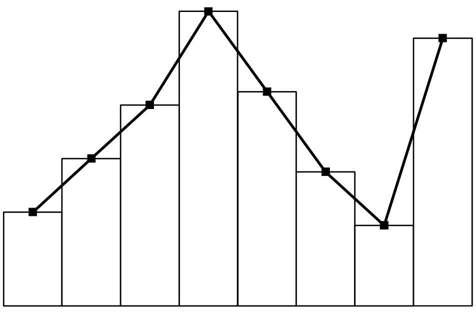


Figure 14h. Edge peaked histogram



8. HISTOGRAM

QUESTIONS

1. How is a histogram different from an ordinary bar chart?

2. What is the formula to use for calculating the range of a histogram?

3. How many bars or categories would be appropriate in a data set of 75 values? 150 values? 225 values?

4. What could be a plausible physical explanation for a double peaked pattern in a histogram? For a skewed pattern?

5. What is the recommended minimum number of values needed to constructed a histogram? Why?

6 What alternative tool is recommended if there is insufficient data values to construct the histogram?

See recommended solutions on page 61.

9. BOX PLOT

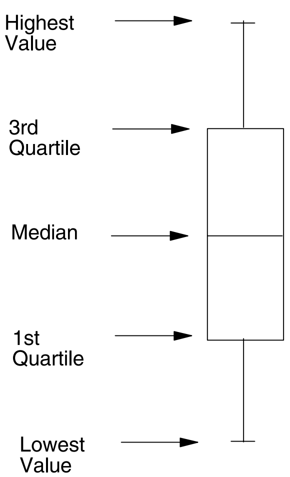
CONCEPTS

Box plots, like histograms, provide a graphic summary of the pattern of variation in a set of data. Box plots are especially useful when working with small sets of data or when comparing many different distributions. We will focus on the basic box plot, although box plots can be used in a variety of ways depending upon the application.

The basic box plot is a five-number summary of a set of data. These five numbers are shown in Figure 15 and are defined as follows:

- Lowest value: the smallest value in the set of data.
- First quartile: the point below which 25% of the data lie.
- Median: the middle value in the set of data.
- Third quartile: the point below which 75% of the data lie.
- Highest value: the largest value in the set of data.

Figure 15. Box plot



The box on a box plot encloses the central 50% of the observations. Its beginning and end are the first and third quartiles, respectively.

The line dividing the box is the median, which divides the data into two equal groups. Half of the observations lie below the median, and half lie above it.

The lowest and highest values are at the end of lines, sometimes called whiskers, extending from the box (for this reason, the box plot is often referred to as a box and whisker chart).

The key benefit of the tool is that the five simple summary statistics, shown in the form of a box plot, enable a team to see some of the aspects of the pattern of variation inherent in the data even if the number of data points is too small to produce a meaningful histogram.

EXAMPLE

The administration manager of a large office received complaints about the availability (or rather, the unavailability) of copier machines. Investigations revealed that while users felt that the current two machines in the building were adequate enough, the copiers were frequently out of service, awaiting the arrival of the repair person. Further investigation revealed that failure rates of the copiers were within reasonable bounds. The response time of the repair person was the major controllable factor in the availability of the copiers.

The copiers were serviced by two different firms. The QI team formed to investigate this problem decided to conduct a simple study to determine the response of each to a repair call. The team defined "response time" as the time between the call for repair and the arrival of the repair person. They then collected the data by comparing various records and log books on a random sample of visits by each contractor. Due to the difficulty of collecting the data, a small sample of 10 visits was collected.

Figure 16a. Response time in minutes (raw data)

Contractor A			Contractor B		
243	345	207	192	251	156
268	272	290	102	228	215
45	251	226	125	145	279
183			188		

The small samples negated the use of a histogram as an investigative tool in this case. So a box plot was used. The data was reformatted into an ordered set as follows:

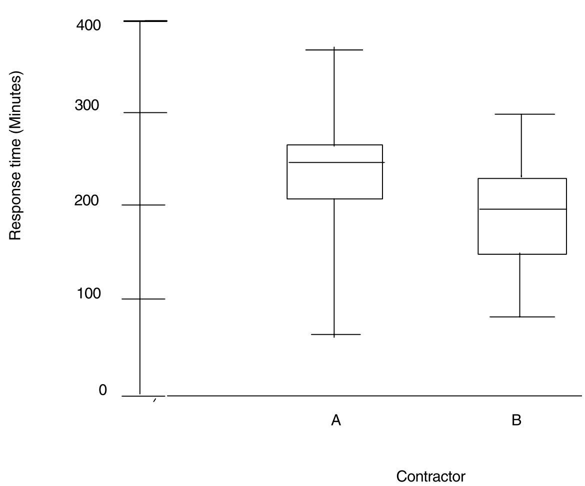
Figure 16b. Response time in minutes (raw data)

Contractor	1	2	3	4	5	6	7	8	9	10
A	45	183	207	226	243	251	268	272	290	345
B	102	125	145	156	188	192	215	228	251	279

The box plot for each contractor was then constructed (see Figure 16c). The team could now draw the following conclusions:

- The median response time of B was 190 minutes, compared to 247 minutes for A.
- The variability in A's response time was greater than B's.
- Even though A had the shortest response time, only about 25% of A's response time were less than 207 minutes, while 75% of B's were less than 228 minutes.
- Nearly 25% of A's response times were longer than B's longest response time.

Figure 16c. Box plot for each contractor



From these observations, the team was able to conclude that B was providing the superior service. The box plot gave them a picture of the limited data available and enabled them to see some aspects of the pattern of variation.

9. BOX PLOT

CONSTRUCTION

1. Collect the raw data and convert it into an ordered data set by arranging the values from the lowest to the highest.

2. Calculate the appropriate numerical values (see formulas below).

3. Draw and label the horizontal axis; provide labels to identify each box plot.

4. Draw and label the vertical axis; label the axis and its unit of measure:
   - begin labeling the axis until you reach a multiple of 1, 2 or 5 that is the smallest value in any box plot.
   - continue labeling the axis until you reach a multiple of 1, 2 or 5 that is larger than the highest value in any plot.

5. Draw the box plots. Include a key to define the various parts of the box plot.

6. Title the chart and show nominal values and limits if applicable.

Definition of symbols:

d = the depth; the number of observations to count from the beginning of the ordered data set
M = the Median or central value
n = the number of observations in the set of data
Q1 = the first quartile
Q3 = the third quartile

Formulae:

Depth of Median = d(M) = (n+1)/2
Depth of First Quartile = d(Q1) = (n+2)/4
Depth of Third Quartile = d(Q3) = (3n+2)/4

9. BOX PLOT

INTERPRETATION

Values in any set of data will vary, and the variation will display some pattern. Box plots are a good way to compare patterns of variation among data sets.

In comparing patterns of variations with box plots, we can look for:

- differences in the location of the median
- differences in the amount of variation
- the presence or absence of outliner points
- symmetry or asymmetry in the data.

Based on this analysis, the team can use its knowledge and further observation to develop theories about the underlying process. In this respect, interpreting box plots is similar to interpreting histograms.

Because they have fewer data points, box plots limit our ability to classify the pattern of variation. The histogram is a much better tool; if you have 40 or more data points, use histograms to analyze the pattern of variation. Figures 17a-h on the following page shows each common histogram pattern and the corresponding box plot shape.

AVOIDING PITFALLS

- Be careful not to use box plots in an inappropriate context.

9. BOX PLOT

Figure 17a. Bell shaped box plot/histogram

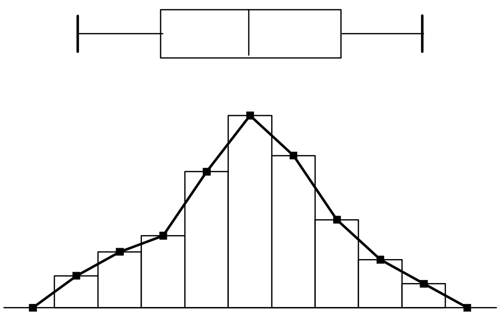


Figure 17b. Double-peaked box plot/histogram

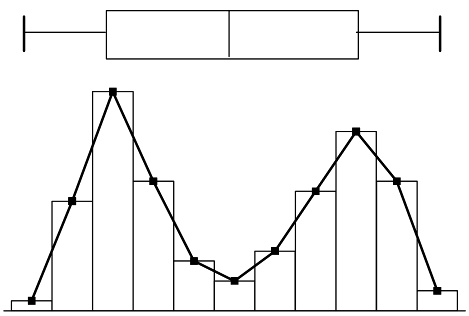


Figure 17c. Plateau box plot/histogram

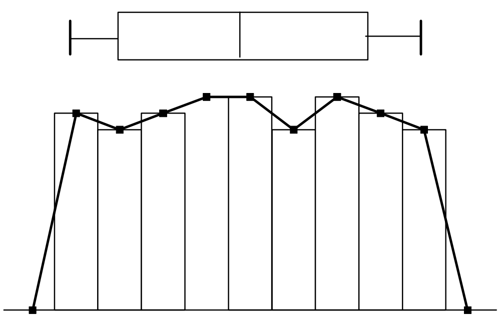


Figure 17d. Comb box plot/histogram

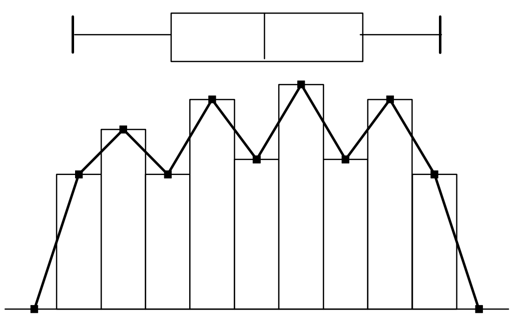


Figure 17e. Skewed box plot/histogram

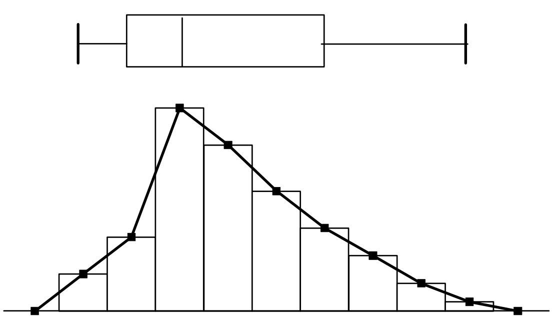


Figure 17f. Truncated box plot/histogram

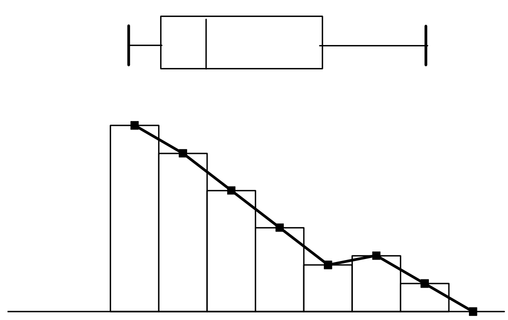


Figure 17g. Isolated peaked box plot/histogram

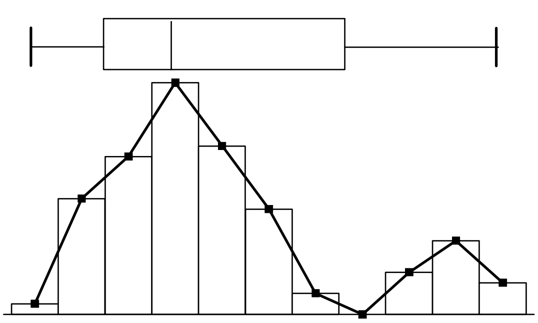
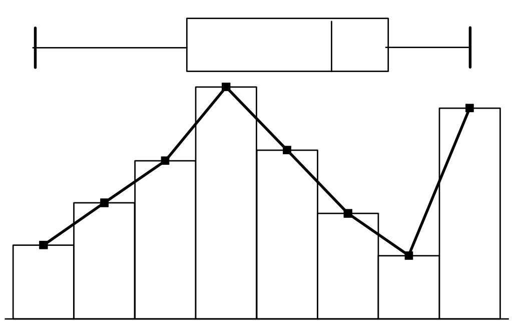


Figure 17h. Edge peaked box plot/histogram



9. BOX PLOT

QUESTIONS

1. When is a box plot considered a good substitute for the histogram?

2. Name the five values plotted in a basic box plot.

3. In a box plot comparing two processes, name at least three features which distinguish the better process.

4. Why is it better to use a histogram rather than a box plot in certain situations?

See recommended solutions on page 61.

10. STRATIFICATION

CONCEPTS

Stratification is the separation of data into categories. It is used most often to identify which categories contribute to the problem being solved. We can classify (or separate) a mass of data into different groups (or categories). Data observations in a given group share common characteristics that define the category. This sorting process is called stratification. Common examples of stratification are stacked bars (vertical stratification) and grouped bars (horizontal stratification).

The characteristic used to separate the data is called a stratification variable. Each stratification variable will have two or more values. A category is defined by specific combinations or ranges of variables and may be defined by more than one variable.

EXAMPLE

A pharmaceutical company learned that one of its previously successful products was gaining a reputation in the medical community as being ineffective. A review of the original clinical studies and current QA lab data offered no insight into the degradation in the performance of the drug. From a brainstorming session, a theory was developed that the active ingredient in the compound could be breaking down more quickly than expected while the medication sat on the shelf. The team decided to investigate this shelf-life theory further.

Past QA lab data indicated that the production process consistently produced products with 149mg to 151mg of active ingredient. The specification called for 148mg to 152mg; so the process was on target and fully capable of producing the product in conformance to specification. Also, past clinical studies showed that a minimum of 120mg of active ingredient was needed for the drug to be effective in use. Shelf-life studies conducted 10 years ago as part of the government approval procedure had shown that the active ingredient broke down at a rate of 1mg/month. So, after the maximum shelf life of 24 months, the drug should still have enough potency to be effective.

Samples of various lots of the drug were pulled from shelves and brought to the lab to determine the amount of active ingredient in mg. The tests included samples that were older than 24 months that have been returned as part of the company's routine procedure for handling expired products. The age of the product was determined from the lot number stamped on each bottle.

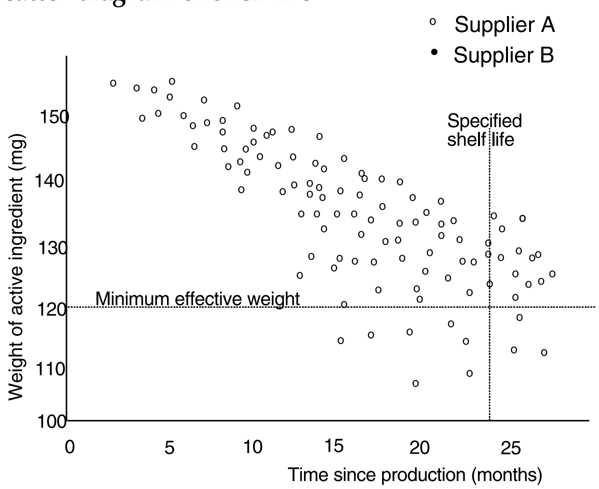
The scatter diagram of the data is shown in Figure 18a. As expected, the graph shows a definite relationship between shelf life and weight of active ingredient. Furthermore, the scatter diagram seems to indicate a non-linear relationship - the breakdown seemed to proceed slowly at first, and faster later on. The key point, of course, is that in many cases the weight of the active ingredient had fallen below the 120mg minimum after as few as 13 months.

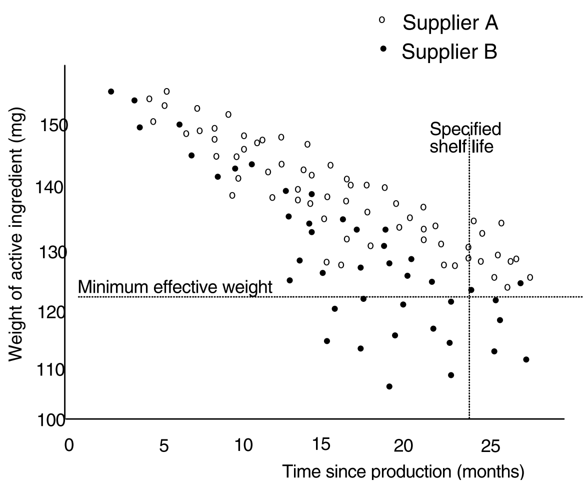
The next phase of the diagnostic journey was to determine the root cause of the poor shelf-life characteristic. Several theories on how the active ingredient may be interacting with some other chemical in the final formulation were tested. By searching through records associated with the various lots of products in the sample, the team finally solved the case by stratifying the scatter diagram it first developed.

Figure 18b shows the same paired data as before, but this time the data had been stratified to reveal something hidden before. A filler ingredient in the formulation of the drug came from two suppliers. In the stratified scatter diagram, lots made with the filler from supplier A are plotted as open dots, while lots made associated with supplier B are plotted as black dots. Two distinct patterns can now be discerned. The lots produced with filler from supplier B showed rapid degradation of the active ingredient, while lots associated with supplier A showed a slower degradation. This combination of degradation rates and differences in the number of data points for A and B accounts for the non-linear relationship that the team observed in the original scatter diagram. Chemical analysis of the filler from A and B revealed the presence of a substance in B's formulation that reacted with the active ingredient to accelerate its breakdown. The remedy involved a change in supplier B's formulation.

Figure 18a. Scatter diagram of shelf life

Figure 18b. Stratification of scatter diagram of shelf life





10. STRATIFICATION

HOW TO STRATIFY

1. Select the stratification variables. If new data are to be collected, include all potential stratification variables.

2. Establish categories that are to be used for each stratification variable. The categories may be either discrete values or ranges of values.

3. Sort observations into the categories of one of the stratification variables. Each category will have a list of the observations that belong to it.

4. Calculate the phenomenon being measured for each category. These calculations are usually either a count of the number of observations in the category or a calculation of an average value for those observations.

5. Display the results graphically for effectiveness.

6. Prepare and display the results for other stratification variables. Repeat steps 2 through 5. Do second stage stratification as appropriate.

7. Plan for additional confirmation. Usually additional data, other techniques, and/or controlled experiments will be used to confirm the initial stratification results.

INTERPRETATION

Looking at the results

If the stratification results are in bar-graph form, it is easy to look across the categories of a variable to see whether one or a few of the categories stand out. Does one supplier have a particularly large defect rate? What types of medication are most prone to errors?

What to do next?

After stratification, if the results give a clear indication of the likely source of the phenomenon being studied, then the team will want to validate their initial results or will want to gain further details on the precise cause.

If the initial stratification does not yield useful results, there are two possible courses of action:

- First, the team could conduct a two-stage stratification that is, stratify by a second variable within each category of a first variable.

- Second, the team could stratify by other variables.

AVOIDING PITFALLS

- Be careful of concluding too much from the data. Small differences among classes should not be given undue weight. If the team is looking at a major quality problem, they should not expect that the causes will be found in small differences.

- Another mistake is to jump to the conclusion that the anomalous category is the cause of the problem. The anomalous category is probably where we should go looking for the cause, but the category itself is not necessarily the cause. If the team is going to collect new data, they should make an effort to collect as much identifying information as they think could possibly prove useful for stratification.

10. STRATIFICATION

QUESTIONS

1. Define the term "stratification" in the context of analyzing data. Which other statistical tool(s) can stratification be used with?

2. Where is stratification most effectively used in a quality improvement or problem solving project?

3. How are stratified variables and their underlying meanings best illustrated?

4. What are the pitfalls to avoid when using stratification to analyze a set of data?

5. Stratification by itself will not lead a team to the root cause of the problem under investigation. What needs to be done next?

See recommended solutions on page 62.

SUMMARY

At this point, you should be able to appreciate the power of data analyzing tools in your problem solving and process improvement projects. Knowing when and how to construct the appropriate diagram is only half the battle. Interpreting the pattern and drawing the right conclusions are the key steps in pointing your team in the right direction towards resolving the problem or improving the process.

In general, all projects must start with a clear and unambiguous mission statement - a clear description of the problem to be addressed and the team's stated objectives when this has been achieved. Existing data from files, logs and historical databases provides an unbiased and easily available source from which the team can establish current performance measures before deciding on targets.

At the outset of the problem solving journey, flow diagrams of the system/process under investigation provide members a common understanding of the "as is" picture. They also provide a means for the team to determine the scope and boundaries of the project. From here, specific areas or sub-processes can be identified as useful points to initiate investigations or data collection efforts. A Pareto of these constituent contributors can then throw up the vital few issues to be resolved immediately.

Brainstorming and cause-effect diagrams are tools used to organize word data. Used consistently, they allow a team to generate useful data in a systematic and focused way. Because several potential issues to be investigated could be generated, a Pareto may once again be used to sieve the vital few from the useful many.

If actual data from the process under investigation is available, line graphs, bar and pie charts are the appropriate tools to use to determine past and current status and trends.

A brainstorming session could again be conducted to generate theories on the suspected causes. New (or existing) data could then be collected to test to prove or disprove these theories. Once again, the team could rely on the flow diagram to identify appropriate data collection points. Alternately, a histogram or box plot could be used to highlight deviations from the normal process. The deviations, interpreted correctly could also point to the root cause of the problem.

Another useful tool for testing new data is the scatter diagram. A strong relationship testifies to the accuracy of the theory put forth; a weak relationship points to an underlying source or problem as the root cause. Even a "no relationship" interpretation assists the team. It says simply that the theory is wrong and can be dropped without using further resources in that direction.

Following this iterative loop of theories testing, the team would have identified the root cause to the problem (or, a few root causes in a process improvement project), and can now collect new data to confirm this (a new scatter diagram or histogram/box plot should show a similar pattern). Having confirmed the root causes to be resolved, corrective action could be planned and implemented. Planning tools (not discussed in this guide) like tree diagrams, barriers and aids charts and implementation tables could be used here.

After implementing corrective actions, new data may be needed to determine how the new/modifed program is performing. Again the various tools used above - line graphs/bar and pie charts, histograms/box plots, scatter diagrams, Pareto charts and so on can be used to monitor improvement and/or progress. Comparisons between before and after diagrams also provide a clear way to show the effectiveness of the remedial actions.

SUMMARY

At any point during the data interpretation/analysis, the team can use stratification as a tool to gain further insights or depth to conclusions already reached with any of the other data analyzing tools, most commonly, bar charts, scatter diagrams and histograms.

Finally, control charts (not discussed in this guide) and other SPC related tools should be used to monitor the process and highlight deviations when they occur.

The tools, as you may have guessed, are just one part to an effective resolution of a problem/process improvement. Work teams must also be equipped with the appropriate methodology to provide the team a framework on their project journey from project identification and objective setting, through the diagnostic journey, problem resolution, planning and implementation, to monitoring and maintaining achievements.

The last element of a successful problem/process improvement project addresses the human factor, or team dynamics. To be effective in a team, members must be trained in and/or made aware of basic skills in team activity. ST University and site training departments offer a variety of courses addressing this topic. See the ST Learning Campus at learning.st.com for more information on available programs.

With all three elements of an effective team, your team can make a positive impact on the environment you work in. Effective interaction between people, strictly adhering to a prescribed and proven methodology and frequent, appropriate application of quality improvement tools is what we, as a company, are striving towards.

SELF-TEST

Each of the questions below indicate a scenario or situation that calls for the use of a particular tool or set of tools. The tools are:

1. Flow Diagram
2. Brainstorming
3. Cause-Effect Diagram
4. Line Graph/Bar Chart
5. Pareto Chart
6. Histogram/Box Plot
7. Scatter Diagram
8. Stratification

Indicate for each case which is the most appropriate tool. (It could be the same tool for several cases.)

1. A new team is meeting for the first time. Members, who work in the same area, are required to identify and commence a project to reduce the number of rejects generated in their area. The team leader wants each member to identify his/her area of responsibility where from experience, there is a (any) likelihood of a reject occurring.

2. After a particularly successful brainstorming session, a group of people listed more than 40 likely causes of time delay in processing a sales order document. It was immediately clear that not all of these were root causes; several were actually the consequence of errors/mistakes created earlier. The team saw the need to separate the data to reflect this distinction.

3. A cross-functional team was formed to resolve increasingly frequent customer complaints about poor response to their queries. The members were representative of the various functions/departments involved. At the first meeting, it was evident that each member was familiar with his/her department's role in the query response process, but not necessarily with what happens before or after that stage of the process. The leader felt that everyone must be familiar with the whole process before tackling the problem.

4. A team of workers was chartered to determine ways in which their dicing machine could process more wafers and thereby increase throughput in their module. One member suggested increasing the speed of the saw for faster dicing and therefore, cut more wafers in a given time. Another more senior operator objected, stating that increasing saw speed would also mean increasing the MTBF (mean time to failure) of the fragile saw, and therefore it could actually slow throughput if machine down time were increased due to the need to change saws more frequently.

   "Fine", said the leader, "let's put this theory to the test. We will dedicate one machine to dicing at an increased speed and collect data on the number of saws used each day".

5. Members of a team investigating ways to reduce cycle time to process a credit note collected a sample of 100 past transactions to determine the average cycle time. It turned out that the average was 20 work days, starting from the time a credit note request was received, to the time a credit note was generated. They also determined that the process involved a total of four departments. The team felt that it would be useful to know the spread of the 100 transactions over the 20 day cycle, as well as how this 20 day cycle was spread over the four departments.

SELF-TEST

6. In the sale order problem stated earlier, the team finally arrived at the conclusion that of the 40 likely causes, 15 were distinct, potential root causes. Due to the limited manpower and time resources, it was not practical to simultaneously investigate all of them. The team had to identify those few that, in their consensus, would bring about the most optimum benefit.

7. In a study on consumption of paper, team members felt it would be useful to know the quantity of new paper consumed throughout the office. However, when approaching the question of where and who is best to collect the data, members were stumped.

8. A team was tasked to reduce the time required to perform a critical but complex task. This task, although essential to the overall operation, was performed only when the item produced failed to meet a particular criterion. Hence, the operation was infrequently performed and data was not easily available. The team undertook some remedial action and wanted to determine the extent of improvement that was been achieved.

9. As part of their SPC efforts, workers in charge of a particular process wanted to determine if the critical measurements in their process conform to the defined control and specification limits. In the absence of the relevant automated data capturing software, they collected data on a daily basis over a period of 30 working days on these critical measures.

10. A problem solving team in the accounts office had the task to resolve a recent and increasingly frequent number of customer complaints about being invoiced for the wrong items. In reviewing 60 erroneous invoices generated from over the past two months, they were able to determine that in each invoice, the error occurred in one of three categories: the day of week when the invoice were generated, the clerk who generated the invoice, and the week of the month in which the invoice was generated. The team wanted to determine which of these categories would be meaningful to investigate.

See recommended solutions on page 63.

RECOMMENDED SOLUTIONS

1. FLOW DIAGRAM

1. When and where can the flow diagram be utilized?
- When a common understanding of the investigated process is needed, without having to walk through the actual, physical process.
- At the diagnostic phase of a project to identify areas for further analysis.
- To identify data collection points.
- To determine the boundaries of a process.

2. What is the symbol for an activity? a decision? a connector?
- A rectangle, rhombus (diamond) and a circle with the appropriate code letter or numeral, respectively.

3. When is a high level diagram best used?
- The high level diagram is an appropriate point to end when the team is mapping the first process flow diagram. This will not unnecessarily clutter up the process picture, thereby allowing the team to identify the problem sub-processes for which the detailed diagram is needed.

4. How do you proceed when you encounter a branch point when drawing a flow diagram of your process?
- Continue with one branch until you reach a suitable terminal point, then return to branch point and continue with the other branch. Repeat this until all branches are addressed. (To facilitate understanding of the process flow, try to draw branch points as a 2-way, yes/no branch-off, with one branch continuing the process and the other addressing other branching options).

5. How can you use an accurate flow diagram to better understand a process?
- An accurate diagram will provide information on the process "as is". To gain a better understanding of the process, users should "walk through" the diagram, step by step and ask appropriate questions addressing issues of poor performance, redundancy and necessity.

6. What are the common pitfalls to avoid when using the flow diagram as a tool?
- Failure to ensure a complete representation of the process will result in the team addressing the wrong issues.
- Using a complex and detailed flow diagram during the early stages of the project may unnecessarily tie down the team in non-critical areas of the process.

2. BRAINSTORMING

1. What is the first step in a proper brainstorming session?
- A clear and unambiguous description of the subject to be brainstormed helps keep participants focused.

2. Name the conceptual and practical rules to observe during a session.
- The conceptual rules: No criticism or evaluation of any kind, be unconventional, aim for quantity of ideas, hitch-hike on other contributions.
- The practical rules: Make contributions in turn, only one contribution/idea per turn, you may "pass" if no contribution is immediately forthcoming, do not provide explanations during the session.

3. How do we handle the large number of ideas generated in a successful session?
- Merge similar ideas and drop those that do not impact the subject matter. Organize the remaining contributions to show up the causal relationships between them. This will help identify potential root causes to some extent.

4. Is there a need to collect data to support ideas/theories generated from brainstorming session? Why?
- Ideas on their own constitute theories. By themselves, they may or may not explain the observed discrepancy/problem. There is therefore a need to collect data to confirm or drop the suspicion.

5. How can a team avoid the dominance of one or a few individuals during a session?
- Team members can avoid this problem by agreeing to strictly adhere to the conceptual and practical rules during the brainstorming session.

3. CAUSE-EFFECT DIAGRAM

1. What is the primary purpose of a cause-effect diagram?
- It is to organize ideas systematically to display causal links between them which lead to the observed effect.

2. In a typical C-E diagram, what description is placed in the box on the extreme right?
- The effect or symptom for which the causes must be identified.

3. Which common tool could a team use to generate inputs for their CE diagram?
- A brainstorming session generates useful inputs for the C-E diagram.

4. Name the 2 checking activities to do immediately after completing the diagram.
- Check for logical validity between each causal chain, and check for completeness of the diagram.

5. What needs to be done after analyzing to CE diagram?
- Results must be substantiated by data collection and analysis.

6. What is the danger of limiting the input of ideas when drawing up the diagram?
- If the ideas put forth are not comprehensive, the real root cause(s) could be masked or ignored, thereby leading the team away from an effective resolution of the problem.

4. DATA COLLECTION

1 Why does data collection constitute an integral part of the problem solving process?

- The proper analysis and interpretation of the appropriate data in a project will keep the team on-track and minimize the danger of utilizing limited resources to mask or partially eliminate the root causes to the problem.

2. What are 5 key considerations when planning to collect data for your project?

- What questions need to be answered?
- How to recognize and communicate these answers?
- What appropriate tools to use to communicate these answers?
- What type of data to collect?
- Where in the process can the data be collected?
- Who in the process can provide the data? who can collect?
- How to minimize errors and/or maximize resources?
- What additional data needed for future analysis or traceability?

3. What are the major considerations when designing a data collection form?

- "KISS" principle; aim for self explanatory designs.
- Reduce opportunities for error.
- Ensure proper references and traceability.
- Incorporate professionalism to solicit interest and cooperation.

4. A common bias to avoid in data collection is Non-response bias. Can you recall what this is?

- Non-response bias is the phenomenon where a significant portion of the target group did not respond to the data collection exercise; eg, the graveyard shift in a 24-hour plant decides not to participate in an opinion survey scheduled immediately after their shift.

5. Discuss the merits and disadvantages of using existing data as opposed to collecting new data.)

- Existing data are (usually) unbiased, easier to collect and compile and take much less time and effort. In contrast, collecting new data requires proper designing of forms, identifying and training collectors. All this will take up significant time and people resources. In addition, care will have to be taken to ensure non-biases on the part of both data sources and collectors.

5. GRAPHS AND CHARTS

1. When is a line graph best used? a bar graph? a pie chart?
- Line graphs are best for showing current status and trends. Bar charts are used when a comparison between categories is necessary and pie charts are frequently used to illustrate proportions contributing to the phenomenon under study.

2. In a line graph, what are the 2 principal axes called? Which one is placed horizontally, which vertically?
- The 2 axes are the independent axis (horizontal) and the dependent axis (vertical).

3. What is the difference between a grouped bar and a stacked bar presentation?
- A grouped bar chart is an example of a bar chart stratified horizontally, showing the contributions of related sub-categories contributing to a particular category. An example is Sales volume by geographical area within country.
- A stacked bar is an example of a vertical stratification. It is particularly useful to show how sub-contributions add up to the whole for each contribution category.

4. What are the key concepts to bear in mind when preparing a graphic chart?
- Graphic integrity, consistent scales and symbols, easy reading and simplicity.

5. How does varying the scale on one axis affect the presentation of a line graph?
- Varying the scale results in a distorted graph. Excluding a significant portion of the proper range from which the data is collected/measured may result in exclusion of data.

6. PARETO ANALYSIS

1.How is a Pareto chart different from the normal bar chart?

- The Pareto chart is a combination of a bar chart and a cumulative line graph. Additionally, the bars in a Pareto chart are ordered or ranked from the tallest on the left to the least or shortest on the right.

2. State the rationale behind the Pareto or 80/20 principle.

- The 80/20 principle is a widely observed phenomenon where many (80%) of the observed effects could be linked to a small group (20%) of attributable causes. As a consequence of this, the principle is a guiding force in problem solving efforts where limited resources could be focused on the vital few causes and result in a maximum return on investment.

3. Name the three elements in a Pareto chart.

- The Contributors, the Contributions and the Cumulative percent of contribution.

4. How can the Pareto tool be used to illustrate improvements in a targeted area?

- A comparison between "before" and "after" charts will show how effective the remedial or corrective actions have been in reducing or eliminating the vital few contributors to the observed problem.

5. When, after constructing a chart, what steps can a team take if it appears that there are no obvious distinctions between the various categories?

- One reason why such a result has been obtained is that the categories of contributors have not been appropriately defined. Try to think of other measures.

7. SCATTER DIAGRAM

1. (What is the primary purpose of a scatter diagram?)
- A scatter diagram is primarily used to test theories i.e., the strength (or weakness) of the relationship between a suspected cause and the observed effect.

2. The variables involved in a scatter diagram are sometimes called the (suspected) cause variable and the (observed) effect variable? Which is to be plotted on the horizontal axis? Which vertical?
- The cause variable is plotted along the horizontal axis; the effect on the vertical axis.

3. Explain the following patterns: a strong relationship? a negative relationship? no relationship?
- A strong relationship clear relationship between the suspected cause and the observed effect; the stronger the relationship, the more accurate the theory put forth.
- A negative relationship shows the observed effect decreasing as the suspected cause is steadily increased. An example is the relationship between the number of hours on duty and alertness: the longer a person is on-station, the more fatigue will set in, thereby decreasing the attention level.
- No relationship indicates that the suspected cause has no observable impact on the observed effect; i.e., the theory proposed is not valid based on observable data.

4. Discuss the danger of interpreting a diagram beyond the range of the data from which it was constructed.
- This phenomenon is also known as data extrapolation. It may be used in some situations, but users should question the validity of this: does the equipment operate beyond the range of the data? is the relationship linear/nonlinear outside the range observed... etc.

5. Scatter diagrams are effective to the extent that they prove or disprove a theory or physical explanation. Where does this theory originate from?
- In the project context, brainstorming sessions and cause-effective diagrams are useful ways to generate such theories. Another source could be from interpretation of the data from other tools - histograms/box plot patterns for example, may indicate variations in the process.

RECOMMENDED SOLUTIONS

8. HISTOGRAM

1. How is a histogram different from the ordinary bar chart?
- The primary objective of the Histogram is to show patterns or variations over a range of data. The horizontal axis of the histogram graph reflects this range.

2. What is the formula to use for calculating the range of a histogram?
- The range is the difference between the maximum and minimum values in the data set.

3. How many bars or categories would be appropriate in a data set of 75 values, 150 values, 225 values?
- Respectively, 7, 8 and 9.

4. What could be a plausible physical explanation for a Double peaked pattern seen in a histogram, a skewed pattern?
- The Double peaked pattern is an indication that 2 similar processes with differing means may be the underlying cause to the observation.
- A Skewed pattern is synonymous with the imposition of a practical limit on one side of the range of measurements, close to the nominal or mean value.

5. What is the recommended minimum number of values needed to constructed a histogram? Why?
- The recommended value is 40 data items. As stated, the objective of a Histogram is to provide a picture pattern inherent in a set of data. Towards this end, the larger the sample, the more confidence we could have that the peak, shape and spread of the histogram is representative of the process under investigation. Although not necessarily precise, the number 40 is a good rule of thumb figure.

6. What alternative tool is recommended if there is insufficient data values to construct the histogram?
- If collection of data samples is difficult or impractical, an alternative to using the Histogram is the Box Plot, which allows for limited visibility of variation in data.

9. BOX PLOT

1. When is a box plot considered a good substitute for the histogram?
- When there is insufficient data available and/or data collection is not practical.

2. Name the 5 values plotted in a basic box plot.
- The maximum and minimum values, the median, the first and third quartile.

3. In a box plot comparing 2 processes, name 3 features to distinguish the better process from the other.
- Difference in the location of the median, difference in the amount of variation (length of each box), the presence or absence of "outliners", symmetry or asymmetry in the data samples.

4. Why is it better to use a histogram rather than a box plot in certain situations?
- Box plots provide a limited view of the pattern of variation in a data sample due to the fewer number of data points.

10. STRATIFICATION

1. Define the term "Stratification" in the context of analyzing data. With which other statistical tool(s) can stratification be used?
   - Stratification is the breaking down of data samples into a second or even third level to gain further insights into the information provided by the first analysis of the data. It is or could be used in conjunction with Bar charts, Scatter diagrams, Histograms... in fact with any other tool if the data available could be broken down into sub-groups or sub-categories.

2. Where is stratification most effectively used in a quality improvement or problem solving project?
   - Stratification is most often used to analyze the observed symptoms. It could also be utilized when testing theories and when attempting to identify a root cause.

3. How are stratified variables and their underlying meanings best illustrated?
   - Stratification is a means to breakdown data into relevant or meaningful groups. Having done this, the next step is to "pictorize" the data by means of one of the other tools, the choice of which would depend on the nature of the data.

4. What are the pitfalls to avoid when using stratification to analyze a set of data?
   - Avoid interpreting too much from the resulting diagram, especially if the differences observed are not distinct.
   - Stratification should be seen as a tool to obtain additional insights to the problem. Do not assume that it will immediately point to a root cause.

5. Stratification by itself will not lead a team to the root cause of the problem under investigation. What needs to be done next?
   - If the results are positive, the next step following a stratification analysis is to collect additional data or details to confirm the theory.

RECOMMENDED SOLUTIONS

SELF-TEST

1. A brainstorming session will allow all members present to contribute likely causes of rejects in their work areas.

2. The Cause-Effect diagram could be drawn to show up the causal links between these likely or potential causes, leading ultimately to the observed problem.

3. A high level or relationship map should be drawn so that all members may see/understand how all the functions relate to one another. A detailed process flow diagram could then be provided by each member responsible for a function.

4. Two theories have been proposed here. In both cases, a scatter diagram could be used to test them. The suspected cause variable is the speed of the dicing machine, while the suspected effects are the number of wafers diced (per unit time) and the number of saws consumed (per unit time).

5. Determining the spread of the 20-day cycle requires a Histogram - the team is contemplating the variation of the process. The idea of breaking up the spread over the various contributing functions is an example of stratification of the available data.

6. A Pareto of data collected on 15 distinct root causes will separate the vital, critical causes from the useful many. This helps the team prioritize the causes in order to address the critical few first.

7. A detailed flow diagram of the operations involved in paper generation will show the team appropriate points in the process to collect the data, as well as ascertaining responsibility for the collection.

8. As data is not readily available or collectable, a box plot should be considered. In this instance, the team wanted to determine the benefits of its remedial effort. A comparison of a box plot before the remedial action and one after should provide the needed analysis.

9. Line graph of the minimum, maximum and mean of the data collected each day superimposed on the control and specification limits would indicate how the process is meeting the requirements.

10. This is another example of stratification, the variables involved being the three defined. Simple bar charts drawn from the data for each of these categories will suffice to show any distortions to the acceptable norm.