EIN 5226

Measure Phase from DMAIC

Part II Introduction

Measurements I

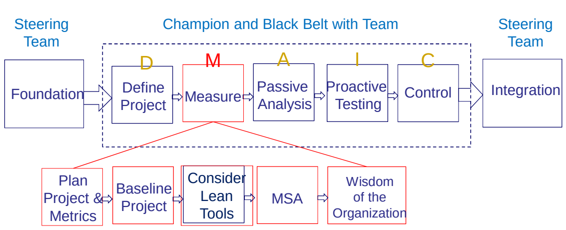
Chapter 3 Sections 1-8

Karen E. Schmahl Ph.D., P.E.

Part II Introduction

- Part II (Chapter 3-14) addresses process definition, process performance, and the quantification of variability.
- KPOVs and KPIVs are identified through consensus.
- Basic analysis tools are introduced:
  - Probability Distributions
  - Six Sigma measures
  - Measurement systems analysis (MSA)
  - Failure mode and effects analysis (FMEA)
  - Quality function deployment (QFD)

Part II Introduction



Chapter 3
Measurements and S^4/IEE Measure Phase

Sections 3.1-8

Chapter 3 Introduction

An objective of the measure phase is the development of a reliable and valid measurement system of the business process identified in the define phase.

- Overview of basic descriptive statistics;
- Data gathering, presentation, and simple statistics;
- Introductory discussion of confidence interval and hypothesis testing;
- Attribute vs. continuous data;
- Ineffectiveness of visual inspections;
- Examples of experiment traps.

3.1 Voice of the Customer

- VOC assessment is needed up front when executing S^4/IEE projects at the 30,000-foot level.
  - Define your customer.
  - Obtain customer wants, needs, and desires.
  - Ensure that focus of project is addressing customer needs.

3.1 Voice of the Customer

- Important customer key process output categories are often classified with regard to their area of impact:
  - Critical to quality (CTQ): flatness, diameter, etc.
  - Critical to delivery (CTD): on-time, accuracy, etc.
  - Critical to cost (CTC)
  - Critical to satisfaction (CTS)
- Important key process input issues are classified as critical to process (CTP)

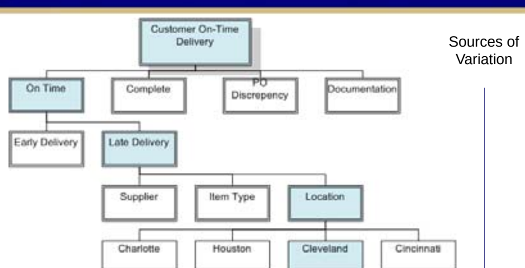
3.1 Voice of the Customer

- The format of a tree diagram can be useful to ensure a linkage between customer requirements and process performance metrics.

  Need→drivers→CTQs

- Tree diagram can also be used to describe the hierarchy of critical to (CT) categories at various level.

3.1 Voice of the Customer: Critical to Satisfaction (CTS) Tree



Cleveland Service Shop Generates Most Defects with a Failure Rate of 89% Late Deliveries

Understanding Variation

- All processes have variation. No two outputs will ever be exactly the same.
- In processes, there are many causes of variation combining to produce an overall effect.
- Variation can be inherent within a system, or can act upon a system to change the system characteristics.

10

Understanding Variation

Types of variation:

- **Common Cause**
  - From a stable system
  - Exhibits random behavior
- **Special Cause**
  - From outside influences

W. Edwards Deming nomenclature (Deming 1986).

11

Understanding Variation

System	Measurement
Driving a car	Gas mileage

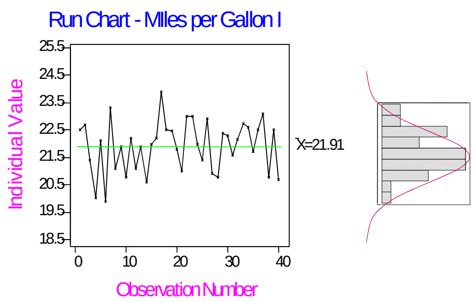
common cause –

What can cause gas mileage to vary with normal use of the car?

special cause -

What might cause an abnormal variation to gas mileage?

12



13

Understanding Variation

Requires understanding of

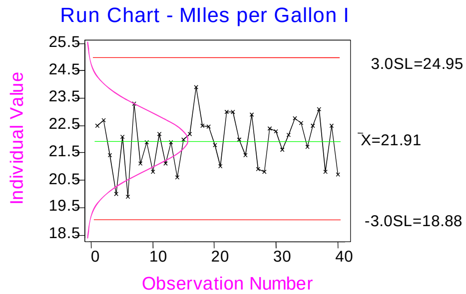
- Response over time
- Central tendency
- Spread of data
- Shape of the distribution

14

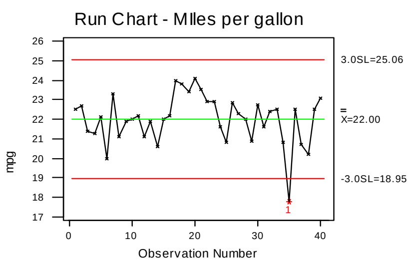
Understanding Variation

- A process that is operating with only common causes of variation present is said to be in statistical control.
  - Random process within identifiable bounds
  - Process stable over time
- Other commonly used terms:
  - Common causes = chance causes
  - Stable system = system in statistical control

15



16



17

3.2 Variability and Process Improvements

- Variability is everywhere.
- Common cause vs. special cause
- Where special causes are identified attempts need to be made to avoid the source of special cause, or prevent it from happening again (assuming negative impact).
- If variability is too large because of common causes, the process needs to be changed.

3.3 Common vs. Special Causes and Chronic vs. Sporadic problems

J. M. Juran (Juran and Gryna 1980) considers the corrective action strategy for sporadic and chronic problems.

- Sporadic problems are defined as unexpected changes in the normal operating level of a process.
- Chronic problems exist when the process is at a long-term unacceptable operating level.

3.3 Common vs. Special Causes and Chronic vs. Sporadic problems

- Process control charts are tools that can be used to distinguish these two types of situations.
- With sporadic problems/special causes the corrective action is to bring the process back to the normal operating level.
- The solution to chronic problems/common cause variation is a change in the normal operating level of the process.
- Solving these two types of problems involves different basic approaches.

3.3 Common vs. Special Causes and Chronic vs. Sporadic problems

The Juran's control sequence (Juran and Gryna 1980) is basically a feed-back loop that involves the following steps:

1. Choose the control subject (i.e., what we intend to regulate).
2. Choose a unit of measure.
3. Set a standard or goal for the control subject.
4. Choose a sensing device that can measure the control subject in terms of unit of measure.
5. Measure actual performance.
6. Interpret the difference between the actual and standard.
7. Take action, if any is needed, on the difference.

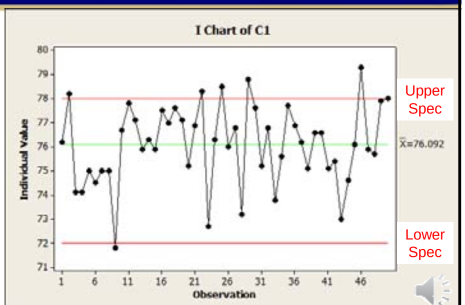
3.3 Common vs. Special Causes and Chronic vs. Sporadic problems

Breakthrough sequence for solving chronic problems

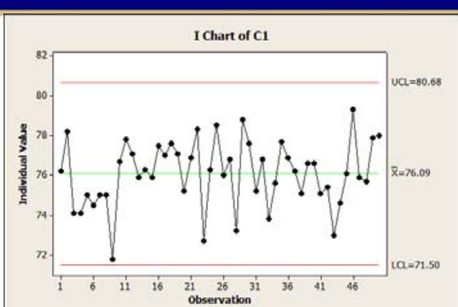
1. Convince those responsible that a change in quality level is desirable and feasible.
2. Identify the vital few projects; that is, determine which quality problem areas are most important.
3. Organize for breakthrough in knowledge; that is, define the organization mechanisms for obtaining missing knowledge.
4. Conduct the analysis; that is, collect and analyze the facts that are required and recommend the action needed.
5. Determine the effect of proposed changes on the people involved and find ways to overcome the resistance to change.
6. Take action to institute the changes.
7. Institute controls to hold the new level.

3.4 Reacting to Data

76.2	76.9	75.1
78.2	78.3	75.4
74.1	72.7	73.0
74.1	76.3	74.6
75.0	78.5	76.1
74.5	76.0	79.3
75.0	76.8	75.9
75.0	73.2	75.7
71.8	78.8	77.9
76.7	77.6	78.0
77.8	75.2	
77.1	76.8	
75.9	73.8	
76.3	75.6	
75.9	77.7	
77.5	76.9	
77.0	76.2	
77.6	75.1	
77.1	76.6	
75.2	76.6	



3.4 Reacting to Data



3.5 Sampling

- Population parameters
- Sample statistics
- Simple Random Sampling
- Sampling error
- Confidence interval

Population vs. Sample

Population

– entire group of interest in an analysis

Sample

– subset of items selected from population

- Sample of size n is taken

26

Sampling

Sample

– Subset of n items

– Should represent the population

Simple random sample

– sample chosen by a method in which each collection of population items is equally likely to comprise the sample.

Sampling

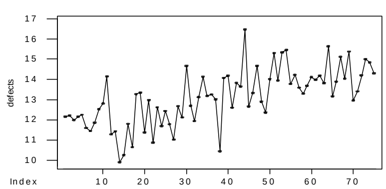
- Samples must come from a well-defined and stable population
  - Determine by initial examination of data
  - If taken over time, must not show cycle or trends

Line Graphs or (Run Charts)

- Generally used to depict time related trends
- Look for stability within system over time prior to doing statistical analysis

29

Run Chart/Time Series Chart



Is this a stable population?
Yes / No
(click on your answer)

30

3.6 Simple Graphic Presentations

- Stem and Leaf Diagram
  - Method of plotting data which displays data values as well as frequency
- Histogram
  - Bar graph displaying frequency of observations in a given bar or interval
- Dot Plot
  - Along a numbered line, a dot plot displays a dot for each observation.

COLLEGE OF ENGINEERING

Graphical Presentations

Graphical displays of data are important tools for investigating samples and populations.

Use to summarize data for easy understanding
- Location or central tendency
- Spread or variability
- Departure from symmetry, shape
- Identification of "outliers"

Stem-and-leaf Plot

- Method of plotting data which displays data values as well as frequency
- Each item in the sample is divided into two parts: a **stem**, consisting of the leftmost one or two digits, and the **leaf**, which consists of the next digits.

33

Stem & Leaf Plot

Sample observations

79	82
109	59
91	102
100	78
75	86
93	85
89	65
63	77
69	102
121	78
74	92
77	73
82	95

Ordered Data

59	82
63	85
65	86
69	89
73	91
74	92
75	93
77	95
77	100
78	102
78	102
79	109
82	121

Stem and Leaf Plot

“Stem”	“Leaf”
5	9
6	359
7	34577889
8	22569
9	1235
10	0229
11	
12	1

34

Frequency Table

Summarizes the data into groupings that show the frequency of data in each group

Stem & Leaf

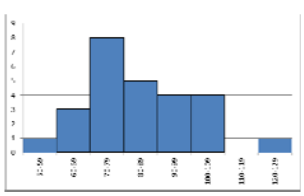
5	9
6	359
7	34577889
8	22569
9	1235
10	0229
11	
12	1

Class Interval	Frequency	Relative Frequency
50-59	1	0.0385
60-69	3	0.1154
70-79	8	0.3077
80-89	5	0.1923
90-99	4	0.1538
100-109	4	0.1538
110-119	0	0.0000
120-129	1	0.0385

Histograms

Stem & Leaf

5	9
6	359
7	34577889
8	22569
9	1235
10	0229
11	
12	1



A histogram is a graphical summary of the frequency of observations in a set of data placed into defined intervals.

Creating a Histogram

- Determine the number of classes to use, and construct class intervals of equal width.
  - Rule of thumb – at least 5 and no more than 15
  - Larger sample sizes -typically have more intervals
- Compute the frequency and relative frequency for each class.
- Draw a rectangle for each class. The heights of the rectangles may be set equal to the frequencies or to the relative frequencies

Creating a Histogram

- Determine the number of classes to use, and construct class intervals of equal width.
  - Rule of thumb – at least 5 and no more than 15
  - Larger sample sizes -typically have more intervals
- Compute the frequency and relative frequency for each class.
- Draw a rectangle for each class. The heights of the rectangles may be set equal to the frequencies or to the relative frequencies

True or False: The chart/bars will look exactly the same whether you use the frequency or relative frequency.

Creating a histogram

Gather data

52	47	54	55	49	51	51	50	51	52	49	49	49	48	52
47	49	49	50	51	51	58	51	47	47	49	48	50	49	49
45	49	47	52	46	48	49	52	47	47	50	48	46	57	54
46	47	51	50	44	48	54	55	53	51	46	53	50	53	49

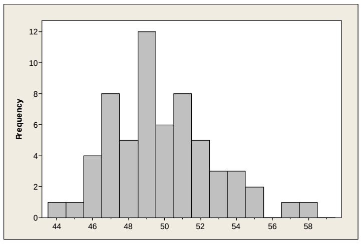
Determine interval size

Often helpful to calculate Descriptive Statistics:

N	Mean	StDev	Median	Min	Max	Range
60	49.850	2.893	49.000	44	58	14

Logical interval options: 1 (14 classes) or 2 (7 classes)

Histogram with Interval=1

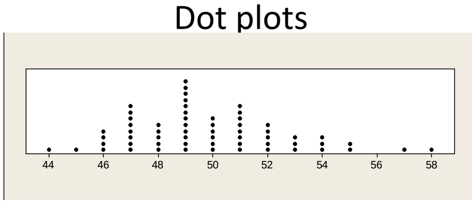


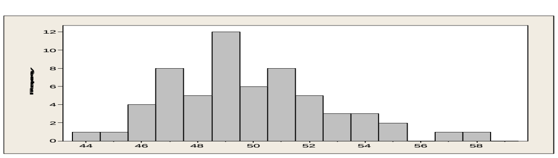
Dotplot

- Along a numbered line, a dot plot displays a dot for each observation
- It is useful when the sample size is not too large and when the sample contains some repeated values.
- Not generally used in formal presentations.

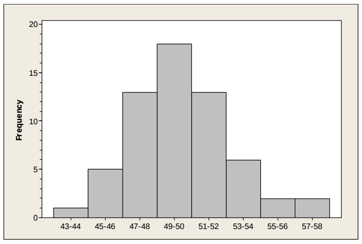
41

Dot plots





Histogram with Interval = 2



3.8 Sample Statistics

Measures of central tendency

- Sample Mean ($\bar{x}$)
- Sample Median (x_{50}, $\tilde{x}$)
- Mode

Measures of dispersion

- Range (R)
- Standard Deviation (s)
- Variance (s^2)

Sample statistics vs. population parameters

Population – entire group of interest in an analysis

Parameter – descriptive number calculated from entire populations values

Sample – subset of items selected from population

Statistic – any descriptive value calculated from the sample group's observations

45

The Arithmetic Mean

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

Where x_i = value of the i^{th} individual observation

n = number of observations in sample

Population Mean = $u = \frac{\sum_{N}^{i=1} x_i}{N}$

Where x_i = value of the i^{th} individual observation

N = number of observations in population

46

Mean

Observation i	Value x
1	13
2	21
3	24
4	12
5	12
6	15
7	19
8	11
9	13
10	13
11	18
12	17
sum	188

- Calculate the mean of the sample

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

$$\bar{x} = \frac{188}{12} = 15.67$$

Median

Median – middle most value

- Arrange data order from smallest to largest, count to middle number
- If even number, average the two middle numbers

Example:

Previous set of data is arranged in ascending order

Average 2 middle value

$$\bar{x} = \frac{13 + 15}{2} = 14$$

11
12
12
13
13
13
15
17
18
19
21
24

48

Median using the Stem and Leaf

Median – middle most value

- Arrange data order from smallest to largest, count to middle number
- If even number, average the two middle numbers

In the stem and leaf plot, we have a set of ordered data.

5	9
6	359
7	34577889
8	22569
9	1235
10	0229
11	
12	1

For this data set, where n=26, what is the median?

A. 78.5 B. 79 C. 80.5 D. 82

Quartiles

50% Quartile – Median

25% Quartile –

Order data
Compute value .25 (n+1)
Count to this observation
If between observations, average the two values

Example - .25 (12+1) = 3.25

$$\bar{x} = \frac{12+13}{2} = 12.5$$

order	
1	11
2	12
3	12
4	13
5	13
6	13
7	15
8	17
9	18
10	19
11	21
12	24

25%
12.5

50%
14

50

Percentiles

50% Quartile = Median
= 50th percentile

pth percentile

Order data
Compute value (p/100) (n+1)
Count to this observation
If between observations, average the two values

11
12
12
13
13
13
15
17
18
19
21
24

25%
12.5

50%
14

51

Percentiles

p^{th} percentile

Order data
Compute value (p/100) (n+1)
Count to this observation
If between observations, average the two values

What is the 60th percentile?

A. 15.5 B. 16 C. 16.5
D. 17 E. 17.5 F. 18

order	
1	11
2	12
3	12
4	13
5	13
6	13
7	15
8	17
9	18
10	19
11	21
12	24

52

Mean or Median

Housing Market for Miami

Average listing price for homes: $556,568
(Week ending Jun 26 2013, 8094 homes for sale)

Median Sales Price: $160,000
(March 13-May 13, 9289 recently sold homes)

Which is more reflective of the cost of housing?

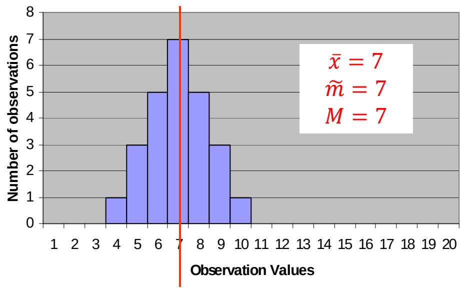
Mode

Mode
– most frequent observation
- may also be used with non-numeric data
- may not exist or may have multiple

11	
12	
12	
13	Mode
13	Mode
13	Mode
15	
17	
18	
19	
21	
24	

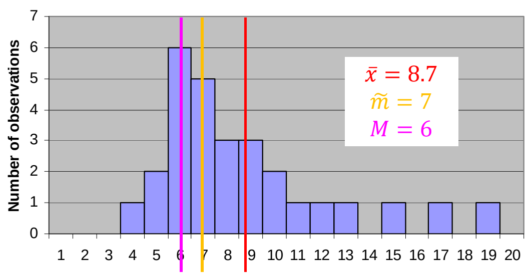
54

Shape of the distribution



With a symmetrical distribution,
the mean, medium and mode are the same

55

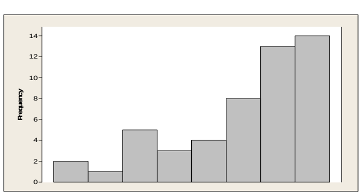


Skewed data – non-symmetrical - more data to one side of mean

Skewed to the right or positively skewed – more data on the right
Mean > Medium

56

Skewed Distributions



The above distribution is skewed left, or negatively skewed.

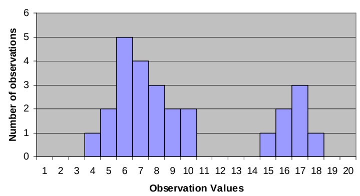
For this distribution,

A. The mean will be greater than the medium
B. The mean will be less than the medium

Mode – most frequent observation

Unimodal histogram– has only one peak

Bimodal histogram – has two clearly distinct modes



58

15.4 Box Plot

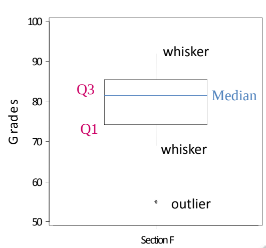
- A box plot (or box-and-whisker plot) is useful for describing various aspects of data pictorially.
- Box plots can visually show differences between characteristics of a data set.
- Common characteristics of a box plot:
  - the lower and upper quartiles (25th and 75th percentiles),
  - the median (the 50th percentile),
  - the interquartile range (IQR),
  - the minimum and maximum within 1.5 IQR of quartiles,
  - and the outliers.

Boxplots

Boxplot - graphical display of important quantitative information about a data set.

Data Set

15	92	
14	92	
13	87	
12	85	Q3= 85
11	85	
10	85	
9	83	
8	82	Median = 82
7	81	
6	81	
5	80	
4	75	Q1=75
3	72	
2	69	
1	55	



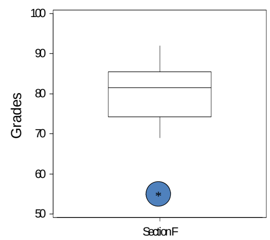
The Box Plot

Outliers - points beyond a distance of 1.5 (Q3-Q1) further than the 1st and 3rd quartiles

Data Set
92
92
87
85
85
85
83
82
81
81
80
75
72
69

Q3=85, Q1=75
Interquartile range =
Q3-Q1=10

Outlier limit – low end
Q1 - 1.5 (Q3-Q1)
= 75-1.5(10)
= 75-15 = 60



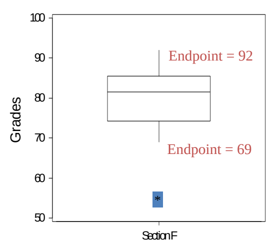
55 Outlier

The Box Plot

Outliers - points beyond a distance of 1.5 (Q3-Q1)

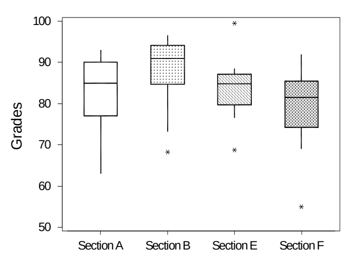
Whisker end points - min and max points without outliers

Data Set
92 Endpoint
92
87
85
85
85
83
82
81
81
80
75
72
69 Endpoint



55 Outlier

Comparison of grades for different sections



Boxplot Comprehension Questions

Answer the following questions using the box plot provided.

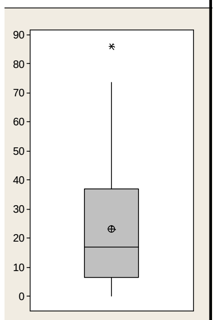
T / F The first quartile is at approximately 37.

T / F Interquartile range is approximately 30.

T / F There are no outliers in the data.

T / F The upper whisker is a bit longer than the lower one, indicating that the data has a slightly longer upper tail than lower tail.

T / F The boxplot suggests that the data are skewed to the left.



Measures of Variation

- Range = Largest – smallest
- Standard Deviation and Variance
  - Essentially a measure of the average difference of value in the sample from the sample mean
  - Formulas vary for sample and population
  - Standard deviation is the square of the variance

65

Standard Deviation

Sample:	standard deviation
	$s = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}}$

66

x	$x-\bar{x}$	$(x-\bar{x})^2$
13	-2.67	7.11
21	5.33	28.44
24	8.33	69.44
12	-3.67	13.44
12	-3.67	13.44
15	-0.67	0.44
19	3.33	11.11
11	-4.67	21.78
13	-2.67	7.11
13	-2.67	7.11
18	2.33	5.44
17	1.33	1.78
188	0.00	186.67

$$s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$$

n=12

$$\bar{x} = \frac{\sum x}{n} = \frac{188}{12} = 15.67$$

$$s = \sqrt{\frac{186.67}{12-1}} = 4.12$$

Standard Deviation

Population: standard deviation

$$\sigma = \sqrt{\frac{\sum (x-\mu)^2}{N}}$$

Sample: standard deviation

$$s = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}}$$

Standard Deviation and Variance

Population: standard deviation

$$\sigma = \sqrt{\frac{\sum (x-\mu)^2}{N}}$$

variance

$$\sigma^2 = \frac{\sum (x-\mu)^2}{N}$$

Sample: standard deviation

$$s = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}}$$

variance

$$s^2 = \frac{\sum (x-\bar{x})^2}{n-1}$$

69

Related Assignments

Please see Blackboard for related assignments.