

Rational Data Analysis—Getting the Best Return

Gaither F. Pennington

1

Session Objectives

- ▶ Describe the impact that small numbers may have on data analysis and successfully identify small numbers in a data set
- ▶ Discuss the appropriate use of control charts, comparison charts, scatter plots and other graphics in the analysis of performance data
- ▶ Describe a rational process of data analysis for performance data

2

Topic Areas

- ▶ Background and theory
 - Process control using statistical methods
 - Graphing approaches
 - Run charts
 - Control charts
 - Comparison charts
 - Scatter plots, histograms, box plots
- ▶ Small numbers, outliers and “power”
 - Practical considerations for proportion measures

3

Topic Areas

- ▶ Process capability
 - Estimating upper and lower process capability limits for continuous data
 - Estimating the effect of common cause variation on continuous data
- ▶ A rational approach to data analysis
- ▶ Two Practical Exercises
 - Control charts
 - Small numbers, outlier identification and process capability

4

A Few Syllogisms...

5

Syllogism #1

- ▶ Major Premise: A duck is a bird that has webbed feet, a spoon-shaped bill and that quacks.
- ▶ Minor Premise: There is a bird that has webbed feet, a spoon-shaped bill and it quacks.
- ▶ Conclusion: The bird I see is a duck.

6

Syllogism #2

- ▶ Major Premise: The rate for measure “A” reflects the process that produced the data used to calculate the measure’s rate.
- ▶ Minor Premise: Our hospital is dissatisfied with our rate for Measure “A”.
- ▶ Conclusion: Our hospital should be dissatisfied with the process for Measure “A”.

7

***To paraphrase James Carville...
“It’s the process, stupid!”***

8

If you are dissatisfied with your performance for a measure, you need to change the process of care for that measure!

9

Syllogism #3

- ▶ Major Premise: Appropriate analysis must be used to evaluate performance effectively.
- ▶ Minor Premise: There is a chance that my hospital may not use appropriate analysis.
- ▶ Conclusion: There is a chance that my hospital might not evaluate performance effectively.

10

***You can't interpret performance
unless you've analyzed things
appropriately!***

11

Statistical Process Control (SPC)

(Process Control Using Statistical Methods)

12

Statistical Process Control (SPC)

- ▶ Based on concepts developed in the 1930s by Walter Shewhart at Western Electric (Bell Laboratories)
- ▶ Refined by W. Edwards Deming and Joseph Juran.
- ▶ Used extensively in industry and some areas within healthcare
 - Clinical laboratory, radiology, nuclear medicine

13

Fundamental Tenets of SPC

- ▶ Variation can be charted and control limits established
 - Processes within control limits are “under control” and can be allowed to continue
 - Processes outside control limits are out of statistical control and require adjustment
- ▶ Variation takes two forms
 - Common cause variation
 - Special cause variation

14

Fundamental Tenets of SPC

- ▶ Eliminating special cause variation
 - Eliminated by those operating the system
 - Operators study problem and identify source
 - Equipment out of adjustment, deviation from policy, etc.
- ▶ Eliminating common cause variation
 - Eliminated by those managing the system
 - Workers constrained within the processes handed them by management
 - Workers cannot independently affect the common cause

15

Fundamental Tenets of SPC

- ▶ Quality generally does not depend on individual motivation
- ▶ Quality depends on more on~
 - Good system design
 - Consistent long-term direction
 - Adequate training
 - Leadership
 - Ongoing evaluation and effective follow-up

16

Graphing Approaches to Process Control

17

Run Charts

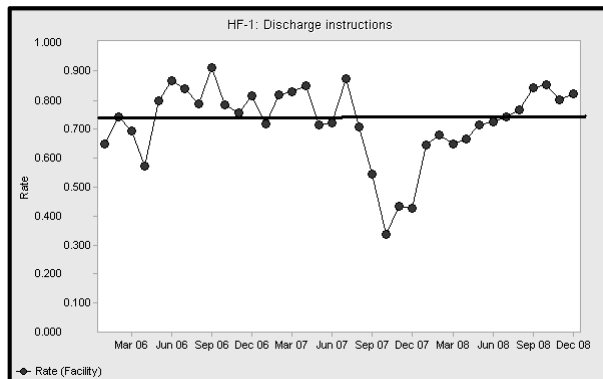
18

Run Charts

- ▶ Advantages
 - Easy to construct
 - Simple to use
 - Don't require assumptions about data distribution
 - Can identify shifts in the process, trends and data oscillations
- ▶ Disadvantage
 - Cannot identify outliers
 - Poor fit for continuous data measures

19

A typical run chart



20

Analyzing a Run Chart

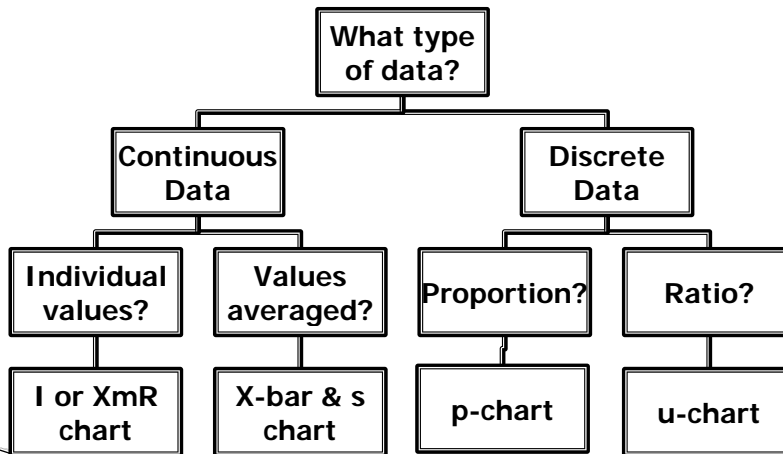
1	How many data runs are there?
2	How many useful data points are there?
3	Are there too many data points in any data run?
4	Is there a series of data points that steadily increase or decrease?
5	Is there a sequence of consecutive data points that alternate up and down?

21

Control Charts

22

Overview for Control Charts



23

Why Create a Control Chart?

▶ ***"Know thy reason..."*** From the Oracle of Gaither

- Reason 1: For statistical uniformity

- One rule -- "Outliers"

or

- Reason 2: For performance improvement

- All rules -- "Outliers", "Shifts", "Trends" and "Oscillations or Cyclical Patterns"

24

Control Chart for Proportions

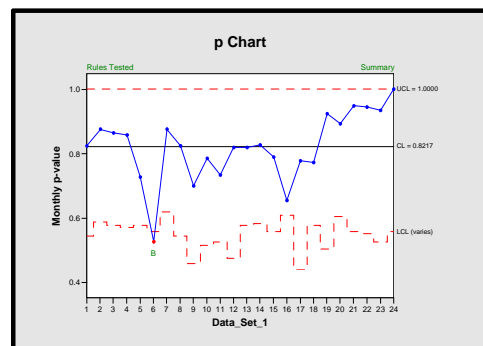
- ▶ Numerator is a distinct subset of the denominator
- ▶ Plots monthly p-values
 - Range between 0.0000 and 1.0000
 - Reported with 4 place decimal precision
- ▶ Graph of choice = p-chart

$$\frac{\text{Numerator}}{\text{Denominator}} = \text{p-value}$$

25

Control Chart for Proportions

- ▶ The p-chart used for all proportion measures
- ▶ Standard ± 3 sigma control limits
- ▶ Varying control limits since subgroups vary from month to month
- ▶ Centerline = process average (mean for the period graphed)



26

Control Chart for Continuous Data

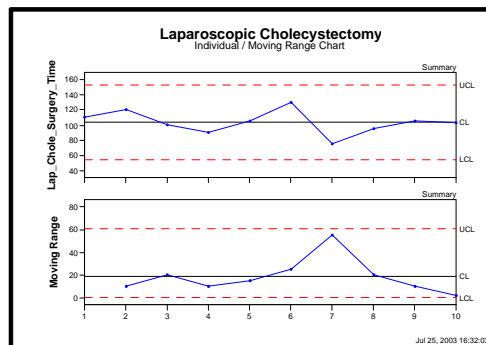
- ▶ No numerator or denominator
- ▶ Values arranged in time-order sequence
- ▶ Reported with 2 place decimal precision
- ▶ Graphs of choice =
 - XmR chart (I-chart) (individual values)
 - X-bar and s chart (grouped values)

(Arrival Time) - (Intervention Time) = Time to treatment (in Minutes)

27

Control Chart for Continuous Data

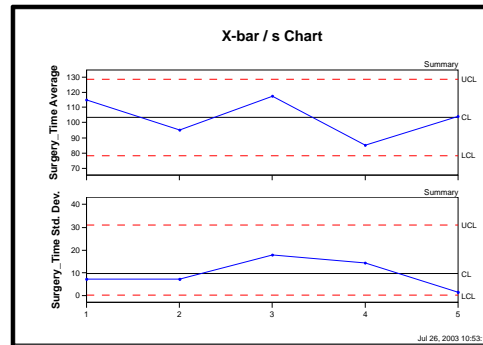
- ▶ The I-Chart or XmR Chart
- ▶ Two charts:
 - Lower = mR
 - Upper = X (individual or "X" values)
- ▶ Standard ± 3 sigma control limits
- ▶ Centerline = process average (mean for the period graphed)
- ▶ Examine lower chart first!



28

Control Chart for Continuous Data

- ▶ X-bar and s Chart
- ▶ Two charts:
 - Lower = s (σ)
 - Upper = X-bar (average of "X")
- ▶ Standard ± 3 sigma control limits
- ▶ Centerline = process average (mean for the period graphed)
- ▶ Examine lower chart first!



29

A Word About Control Chart Rules...

1. More rules = greater likelihood one or more will be broken.
2. Nelson's rules most commonly used in industry.
3. The Joint Commission uses the following rules:
 - a) Outliers
 - b) Shift in the process
 - c) Trends
 - d) Data oscillations or cyclical patterns

30

Common Control Chart Rules

Rule	Definition
Outliers	Any data point above +3-sigma
	Any data point below -3-sigma
Shift in process	≥ 9 data points in a row above the CL
	≥ 9 data points in a row below the CL
Trend	≥ 6 data points in a row, including end points, progressively increasing or decreasing
Oscillations	≥ 14 data points in a row alternating up and down

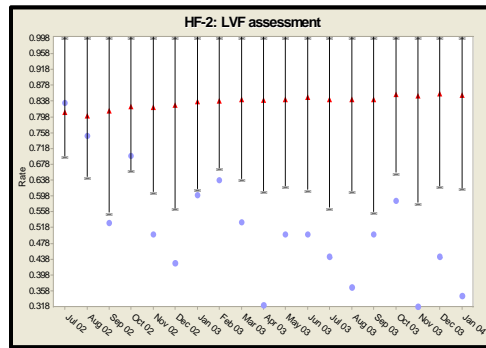
31

Comparison Charts, Box Plots, Histograms and Probability Plots

32

The Comparison Chart

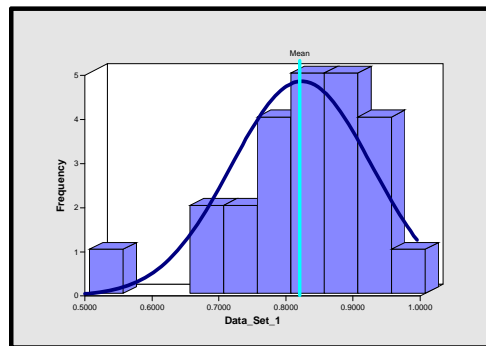
- ▶ The “conjoined twin” of the control chart
- ▶ Places facility in context with others
 - 99% expected performance range [“whiskers”]
 - Facility rate [·]
 - Comparison group mean [▲]



33

The Histogram

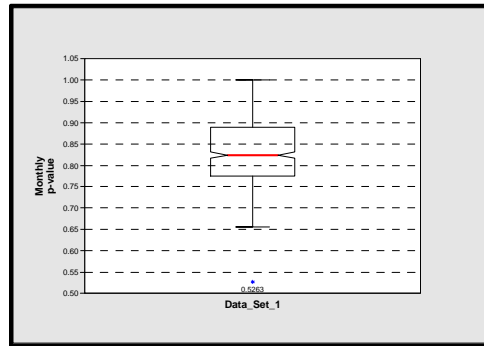
- ▶ Used to chart relative frequency distributions for continuous data
- ▶ Height of bar equal to class frequency
 - Curve and mean reference line are optional



34

The Box Plot

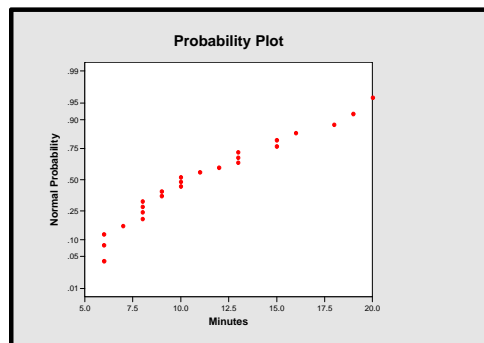
- ▶ Summarizes data in terms of its various percentile values
 - 1st percentile is bottom of box
 - 2nd percentile (median) is at notch
 - 3rd percentile is top of box
- ▶ Outliers /extreme values above or below whiskers



35

The Probability Plot

- ▶ Used for continuous data measures only
- ▶ Determines if data are normally or near normally distributed
- ▶ Do data points pass the “fat pen” test?
 - Data sets failing “fat pen” test should be transformed



36

Small Numbers, Outliers and “Power”

37

Subgroup (Denominator) Size

- ▶ Analysis of data is affected by sample size (subgroup or denominator size)
 - A larger subgroup = more robust analysis
 - A larger subgroup = greater confidence in analysis
 - There is a “critical mass” for analysis
 - Descriptive and inferential statistics
 - Statistical Process Control (SPC)

38

Subgroup (Denominator) Size

- ▶ Subgroups (denominators) are algorithm-defined, and therefore, limited
 - You have to “*dance with what the algorithm brings you...*”
- ▶ There is such a thing as “small numbers”
 - Bad news: Small numbers complicate interpretation
 - Good news: Small numbers can be identified rather than “intuited”

39

The “Power” of a Subgroup

- ▶ A statistic derived from the ***entire data set*** helps determine if any subgroup (denominator) “***lacks sufficient power***” for Statistical Process Control analysis
- ▶ That statistic is called “p-bar”
 - “Bar” in statistics means “average”
- ▶ “p-bar” is the “average of the p-values in the data set”
 - Numerator / Denominator = p-value

40

The "Power" of a Subgroup

When "p-bar" is	<u>Each subgroup</u> (denominator) must be greater than or equal to	If not?
≤ 0.5000	$\frac{1}{p\text{-bar}}$	"Small number" = caution with analysis
> 0.5000	$\frac{1}{(1 - p\text{-bar})}$	"Small number" = caution with analysis

41

The "Power" of a Subgroup

When "p-bar" is	<u>Each subgroup</u> (denominator) must be greater than or equal to	If not?
≤ 0.5000	$\frac{4}{p\text{-bar}}$	Caution if that subgroup is an outlier!
> 0.5000	$\frac{4}{(1 - p\text{-bar})}$	Caution if that subgroup is an outlier!

42

The “Power” of a Subgroup

With insufficient power
relative to:

The options are:

“Small Numbers”

1. Combine subgroups to eliminate the small number
2. Always interpret with caution!

“Outlier Values”

1. Ignore that subgroup
 2. Combine subgroups to eliminate the small number
 3. Always interpret with caution!
-

Small Number and Outlier Exercise

Slide 43

- N1** Well, for the 4/σ rule you can ignore the subgroup. when you are dealing with the 1/σ rule you cannot really do a control chart but need to aggregate up to quarterly or monthly
nmatthes, 8/26/2008

Process Capability for Continuous Data Measures

45

***“Some things you should accept
on faith as long as there is a
good bibliography!”***

46

Estimating Process Capability

- ▶ For continuous data:
 - Statistical theory allows you to:
 - Estimate the Upper Limit of process capability
 - Estimate the Lower Limit of process capability
 - Determine the maximum difference between two consecutive values that can be attributed solely to common cause variation

47

Estimating Process Capability

- ▶ A statistic derived from the data set helps determine estimates of process capability
- ▶ That statistic is the “Median Moving Range (MMR)”
- ▶ To determine Median Moving Range (MMR):
 - Arrange continuous data in time-order
 - Determine the moving ranges for the data
 - Always one fewer Moving Range than the total number of data points
 - Calculate the Median for the Moving Range (or MMR)

48

Using the Median Moving Range (MMR)

Determining the "Dead Band" on either side of the Process Mean

$(\text{MMR} \times 3.14) + \text{Process Mean} =$ Upper Limit Estimate for process capability

$(\text{MMR} \times 3.14) - \text{Process Mean} =$ Lower Limit Estimate for process capability

49

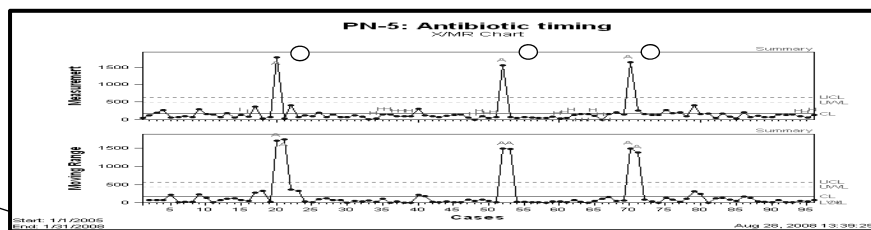
Using the Median Moving Range (MMR)

$(\text{MMR}) \times 3.865 =$ The maximum difference between two consecutive values that can be attributed solely to common cause variation

50

Process Capability Estimates

- ▶ Upper limit estimate of process capability: 315 min
- ▶ Lower limit estimate of process capability: 1 min
- ▶ Max difference due to common cause variation: 218 min
 - 3 cases in dataset > 1,500 minutes!



51

A Rational Approach to Data Analysis

52

Analyzing Proportion Measures

- ▶ Structuring your analysis
 - Verify data accuracy and completeness
 - Use a Proportion Measure Screening Tool
 - Identify denominators that are small numbers
 - Identify denominators that lack sufficient power to interpret outliers
 - Construct and interpret a control chart
 - Correlate chart interpretation with the Screening Tool
 - Construct and interpret a comparison chart

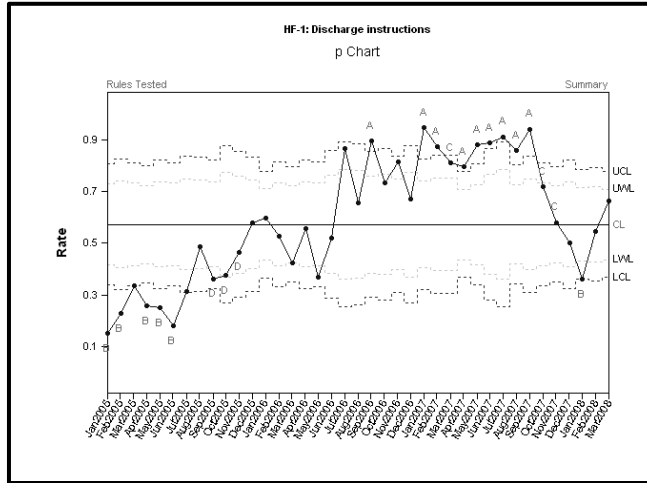
53

Proportion Screening

Measure:		HF-1 Discharge Instructions		Control limits set at ± 3 -sigma, NP = No Population	
		Column A		Column B	
		"Small Numbers"?		"Outliers"?	
		The value of p-bar for this data set is:	Since p-bar is less than or equal to 0.5000, using the formula $1 / p$ -bar, the minimum subgroup size needed for Statistical Process Control (SPC) is	Since p-bar is less than or equal to 0.5000, using the formula $4 / p$ -bar, the minimum subgroup size needed to determine whether data points are outliers is	
$\frac{\text{Total Numerator} = 471}{\text{Total Denominator} = 1,100} = 0.4282$		0.4282	3	10	
Month	p-value	Numerator	Denominator (Subgroup)	When we compare the month's denominator to the calculated minimum subgroup size above, then:	When we compare the month's denominator to the calculated minimum subgroup size above, then:
1 Jan 2004	0.0000	0	1	Small number. Interpret with caution!	Small subgroup. Cannot identify outliers!
2 Feb 2004	1.0000	7	7		Small subgroup. Cannot identify outliers!
3 Mar 2004	▶ 1.1111	10	9		Small subgroup. Cannot identify outliers!

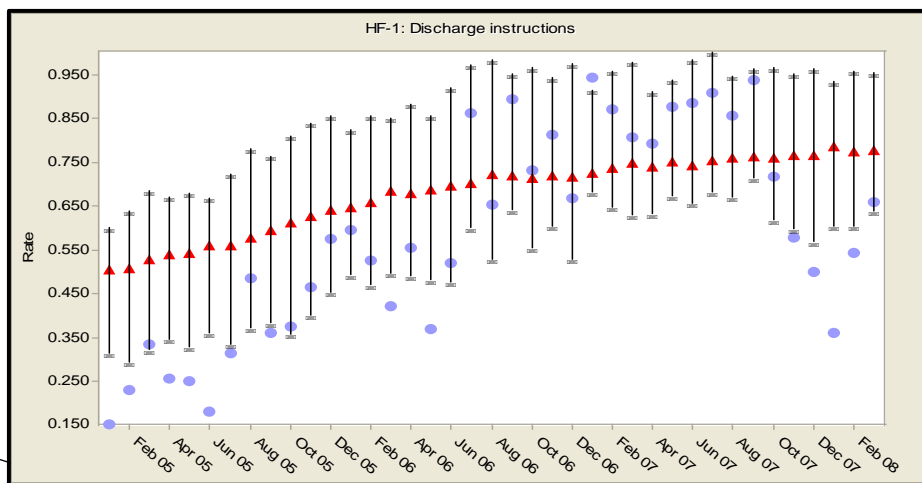
54

The Control Chart: p-chart



55

The Comparison Chart



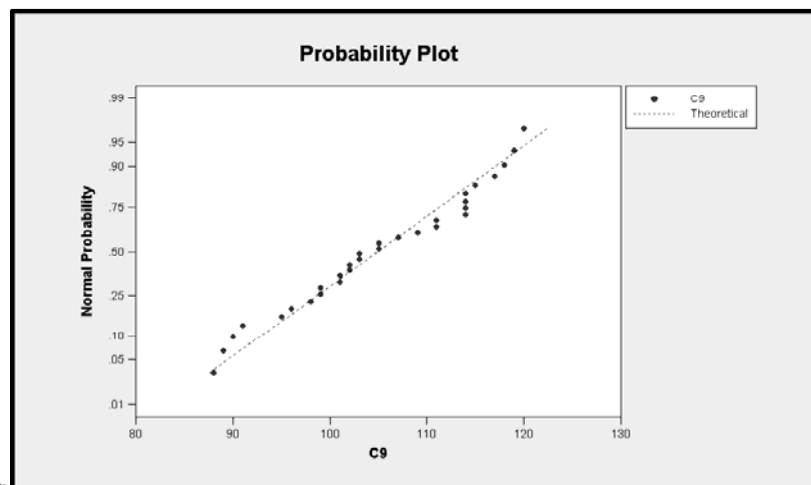
56

Analyzing Continuous Data Measures

- ▶ Structuring your analysis
 - Verify data accuracy and completeness
 - Create a Probability Plot
 - Transform data if it fails the “fat pen” test
 - Create a control chart
 - Create a comparison chart
 - Create additional charts
 - Box plot and/or histogram
 - Estimate process capability

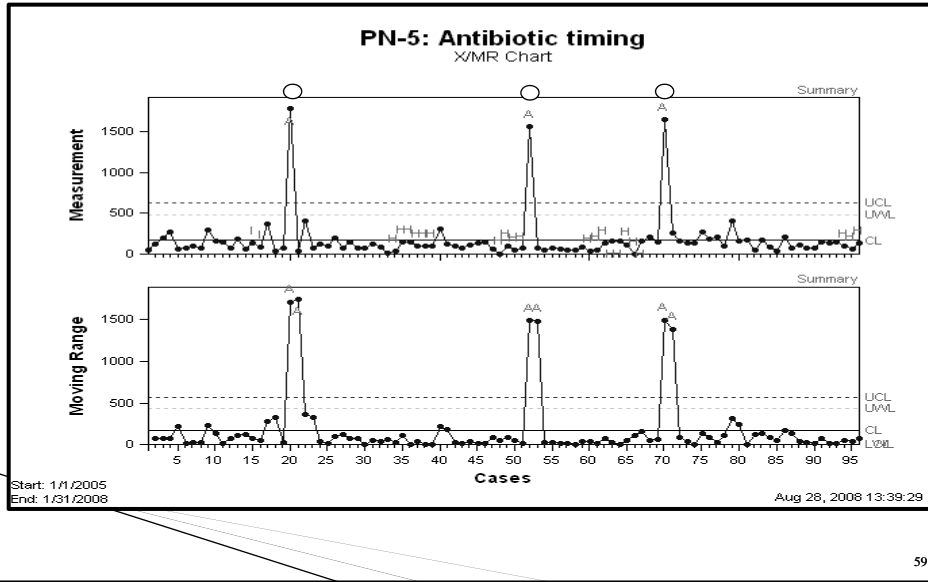
57

The Probability Plot



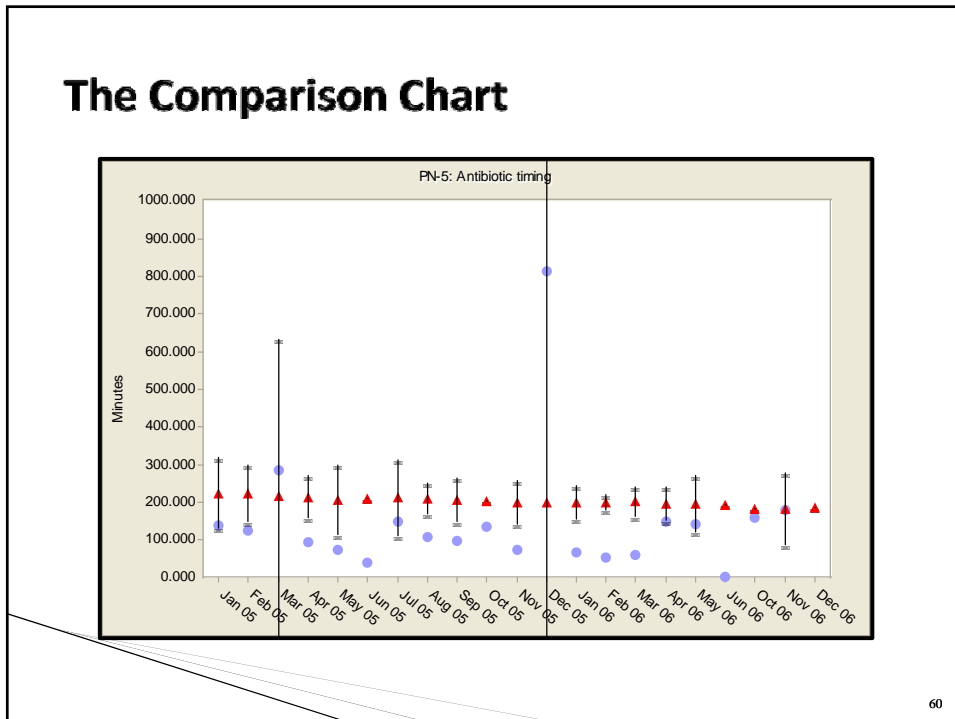
58

The Control Chart



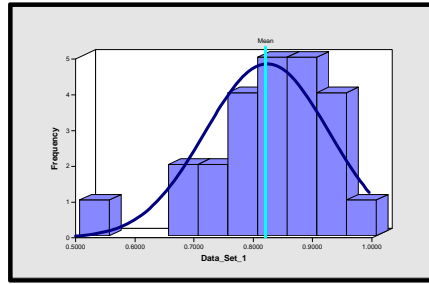
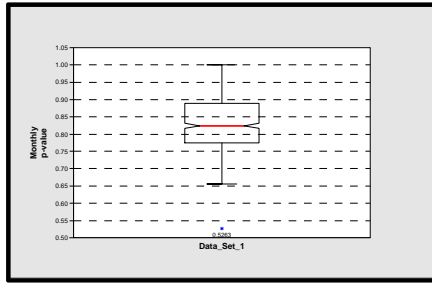
59

The Comparison Chart



60

The Box Plot and The Histogram



61

Putting it All Together Exercises

62

Wrapping it Up... You in the Role of Expert!

63

You as St. John the Baptist?

*A voice
crying in
the
wilderness?*



64

You in the Role of Expert

- ▶ Be honest with yourself about demands in today's performance improvement activities
 - Have realistic expectations for yourself, your staff and facility
 - Have empirical knowledge of resources needed to fulfill your role
 - Be forthright with superiors and colleagues
 - "If you can't stand the heat, you need to get out of the kitchen!"

65

You in the Role of Expert

- ▶ Become comfortable in your role as "expert"
 - Blanche Dubois was right about "illusion"
 - Don't apologize for the data
 - Don't apologize for using new approaches or tools
 - Statistical Process Control (SPC)
 - Process Control Using Statistical Methods
 - Graphing tools
 - Comparison charts, box plots, histograms and probability plots

66

You in the Role of Expert

- ▶ Be certain data are accurate
 - There is no substitute for accuracy and completeness
- ▶ Avoid “spin”
 - Avoid “value” qualifiers like “right”, “wrong”, “good” or “bad”
 - The data are...simply what they are
 - It’s OK not to “like” what you see

67

You in the Role of Expert

- ▶ Understand what you present
 - Command of data elements, definitions and exclusions
 - Know how the algorithm works
 - Become comfortable with:
 - The statistics you report
 - The graphs you report

68

You in the Role of Expert

- ▶ Integrate data with practice and the real world
 - Each Case ID that produces data has a name and mailing address!
 - Focus on “process” not on “persons”
 - Understand the processes of care that produce the data
 - Data without context is worthless
 - Offer solutions, not just “bad news”

69

You in the Role of Expert

- ▶ KISS (Keep It Simple Stupid)
 - Consistency is better than novelty
 - Avoid “audience/data” mismatch
 - The audience changes—not the data or message
 - Beware of “numbers narcosis”
 - Pictures are easier to assimilate than tables of numbers
 - Choose the right picture for the data

70

Questions?

Thanks for Attending!