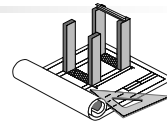


Data and Measurement

“102”

Building for Excellence 2006

November 2006



**Trista Johnson, PhD, Allina Hospitals and Clinics
Director of Performance Measurement and Analysis**

(Attribution Note: The content of some slides came from the Institute for Healthcare Improvement Materials)

Today's Objectives


- Better utilize results published in medical reports and journals by learning about statistical significance, using confidence intervals and p-values
- Describe at a high level the associations between variables, examples of correlation, simple regression, and 2-by-2 tables
- Explain the importance of collecting data over time using run charts and/or control charts
- Explain the distinction between “statistical significance” and “practical significance”



Uses of Data in Health Care

- Research
- Accountability
- Clinical Management/Population Health
- Improvement

3



We don't know the "truth"...
Statistics are used
to estimate the "truth"

4



Quick Review of the Basics

- **Population**= Data from all possible members of a population. (More exact results, but often more labor intensive.)
- **Sample**= Data is from a subset of the proportion and is intended to approximate the population. (Less exact results, but less labor intensive)
- **N**= Sample Size
- **Mean**= The average of responses for a variable in a population (e.g., mean blood pressure in a sample of cases is 110 mm Hg)
Mean = (sum of responses) / N

5



Sampling Methods

In data collection, there are essentially two options:

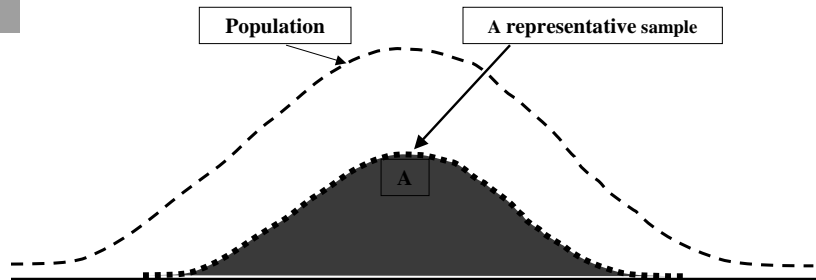
(1) Utilize 100% of relevant available data

OR

(2) Select a sample that, with some luck, mimics 100% of relevant available data

6

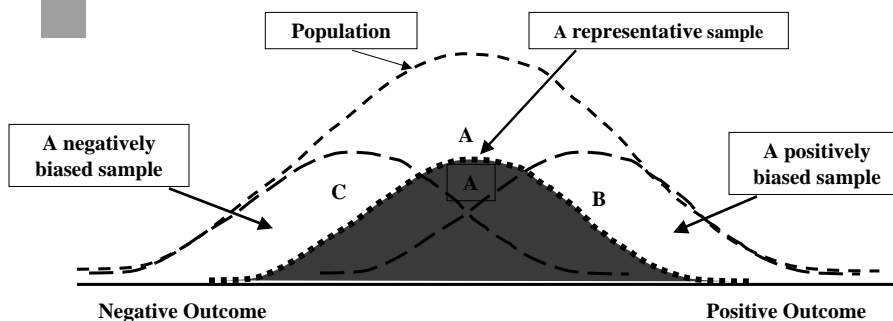
Relation between a Population and a Sample



The “population” is the larger group to whom you wish to generalize the results you obtain by studying the “sample”.

The number of observations in your sample is referred to as the “N”, or “sample size”.

A “Good” Sample



Ideally a “good” sample will have the same shape and location as the total population but have fewer observations (curve A).



Quick Review of the Basics

Sample Size:

- Keep in mind the sample size you will be collecting, as sufficient denominator volumes are required for reliable and meaningful aggregate statistics, such as means.
- General rules: If the population is <10 , aggregate statistics should not be used.
- If the population is 10 to 20, aggregate statistics may be appropriate, but be sure to include a cautionary note about drawing firm conclusions from the data.
- If the population is >20 , there is generally no need for a cautionary note when reporting aggregate statistics.

9



How do you draw a sample?

- Simple random sampling
- Stratified sampling
- Systematic sampling
- Block sampling

Why can't we always just use simple random sampling?

Sometimes random sampling leaves to chance the event that important patterns in the data remain undetected, or worse, dilute or distort results.

10

Stratified Sampling

- Factors known to be related to the response of interest are accounted for during sampling
 - ❖ Example: Randomly select an equal number of observations from the day and afternoon shifts
 - ❖ Allows for depiction of an important and inherent difference in the data that is not related to your improvement strategy
 - ❖ Assures sufficient number of data points from important subgroups

- Facilitates subgroup or “stratum-specific” analysis
 - ❖ i.e., display data separately for day and afternoon shifts

- Disadvantage may be that data collection scheme is slightly more difficult to implement.

Systematic Sampling

- Selection of every k th element from the population

- Collect data at fixed intervals (every hour, every 10th patient...)
 - ❖ Decide on the # of data points you need
 - ❖ Identify or estimate the total observations or units in the time you wish to study
 - ❖ Calculate the interval; divide the total # of units by the number of data points you need

- Example: Billing office sends 1000 bills per week, team wants 100 data points, collect on every 10th bill



Block Sampling

- Decide on number of data points needed
- Select the time and place to begin
- Select the 1st unit at that time and place and collect every unit thereafter until you have the required number of observations
- Example: Start with the first patient admitted to critical care on Tuesday June 13, and collect data on every new patient admitted to the unit until a total of 20 patients is reached.

2004 IHI Information Gathering Tools (On-line)

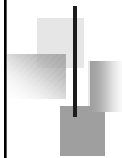
13



Statistical Significance

- Statistical significance measures the likelihood that the result found could have occurred by chance alone
- Used in deductive reasoning: from hypothesis to conclusion
- Should be used in well designed studies

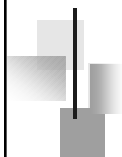
14



Statistical Significance

- P-value: The probability of observing a test statistic at least as extreme as the observed value if the null hypothesis were true
- Null hypothesis: Observations are usually the result of pure chance

15



Statistical Significance

- Hypothesis: Point A is equal to point B ($A = B$)
- Results from a sample show that $A = 2/3B$
- P-value = .05 indicates a 5% chance of finding a difference as great as found ($A = 2/3B$), **IF** A actually does equal B (i.e., the null hypothesis is true)

16



Statistical Significance


- We compare the p-value to a fixed significance level, α .
- α acts as a cut-off point below which we agree that an effect is statistically significant.
- In clinical trials, most results are considered statistically significant if that likelihood is less than one in 20 (or $p < .05$)
- i.e., if $p\text{-value} \leq \alpha$, then we agree something is going on (or that our findings are significant)

17



Confidence Intervals

- Confidence Interval: estimation of the probability of producing an interval containing the actual or “true” value
- 95% Confidence: If you take many samples, the actual value will fall within the confidence interval 95% of the time
- Used in inductive reasoning: reason backward from a set of observations to a reasonable hypothesis

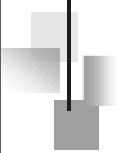


Confidence Intervals

Systolic Blood Pressure Example

- N = 26, Mean = 113.1 mm Hg, SD = 10.3 mm Hg
- 95% CI = mean ± 1.96 SE
 - SE = SD/sqrt(N)
- Result = 109.1, 117.1 mm Hg
- The bigger the n, the smaller the error (i.e., the smaller the interval)

19



Example

Insulin Factors	OR	95% CI	p value
Number of Insulin Orders (continuous)	1.31	(1.22-1.39)	<0.0001
Number of Insulin Types (continuous)	3.08	(2.41-3.95)	<0.0001
Sliding Scale (Yes.No)	1.61	(0.96-2.71)	0.073
Number of Insulin Doses Charged (Continuous)	1.29	(1.15-1.44)	<0.0001

20

Common Types of Variables

- Dichotomous: has only two levels (such as yes/no)
- Categorical: has no measurement scales (food groups, skin color)
- Ordinal: ranked variables not on continuous scale
- Continuous: continuous scale

21

Statistical Analyses

Dependent variables	Independent Variables-> Categorical	Continuous	Mixed
Categorical	Exact tests, Pearson Chi-square, Corrected Chi-square Mantel Test for Trend		
	Mantel-Haenszel Summary Logistic Regression	Logistic Regression	Logistic Regression
Continuous	T-test ANOVA Linear Regression	Linear Regression	Linear Regression ANOVA
Time to Event	Log-Rank Cox Regression	Cox Regression	Cox Regression

22



Correlation/Simple Regression

- Correlation is a measure of how well X predicts values of Y. A scatter plot is a visual description of the correlation
 - This only implies a relationship, not causation
- Regression analysis fits a straight line to a scatter plot.
 - Outliers can skew the fit of a regression line
- X is the independent or predictor variable, Y is the dependent or response variable

23



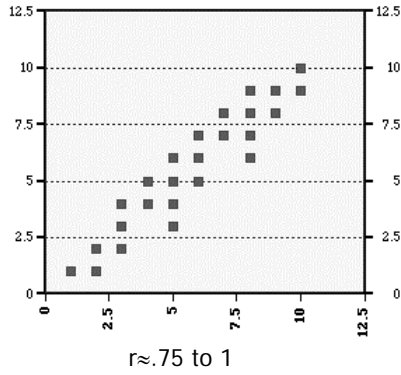
Correlation/ Simple Regression

- The correlation coefficient (r) measures the tightness of fit and whether increasing x relates to a change in y .
- The values of r range in absolute value from 0 to 1. The closer r is to 1, the stronger the relationship between the variables x and y .
- Negative r means that x is **NEGATIVELY** related to y .
- Positive r means that x is **POSITIVELY** related to y .

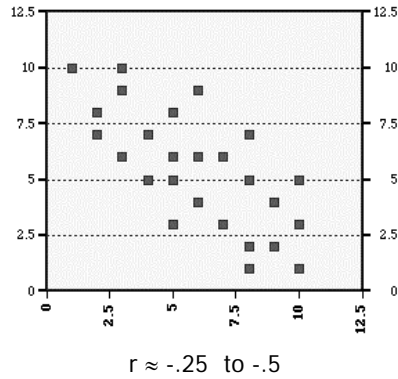
24

Correlation/Scatter Plots

High Positive Correlation



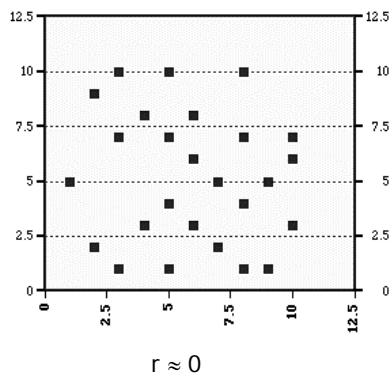
Low Negative Correlation



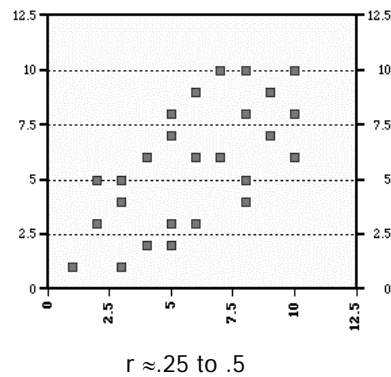
25

Correlation/Scatter Plots

No Correlation



Low Positive Correlation



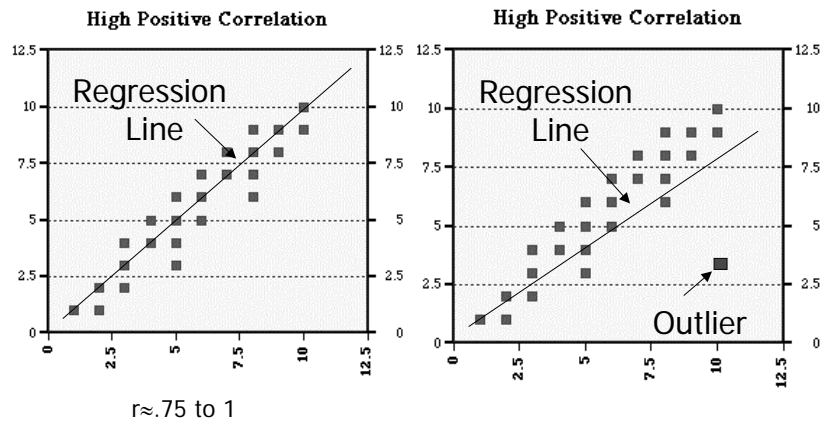
26

Correlation/ Simple Regression

- Examples of when you could use correlation:
 - Satisfaction: what variables are related to overall satisfaction (ex: wait time) Level 4
 - Patient safety: (Nurse/patient ratios vs. safety incidents) Level 3
 - Patient safety: (Nurse/patient ratios vs. safety incidents) Level 1
- Multiple regression is the same as simple regression (correlation), but has more than one predictor variable

27

Regression Lines



28



Two by Two Tables

- Two groups and one 'YES/NO' variable
- Two by two tables can be analyzed using chi-squared tests of independence (i.e., statistical difference between the two groups)
- Chi squared tests need sufficient samples (a rule of thumb: they may be invalid if one or more cells contain a number smaller than 5)

29



Two by Two Table

Outcome	Case Group	Control Group	Total
Exposed	a	b	$a+b=m_1$
Not Exposed	c	d	$c+d=m_0$
Total	$a+c=n_1$	$b+d=n_0$	$a+b+c+d=T$

Used to test if the difference between survival and death is significant based on membership in the treatment vs. control group.

30



Two by Two Tables

■ Chi-square Test:

Measures whether there is a significant difference between the two groups.

$$\text{Chi-square} = \frac{(|a*d - b*c| - T/2)^2}{n1*n0*m1*m0} (T-1)$$

31



Two by Two Table

Outcome	Treatment Group	Control Group	Total
Survived 28 days	38	29	67
Did not Survive	7	17	24
Total	45	46	91

Used to test if the difference between survival and death is significant based on membership in the treatment vs. control group.

32



Two By Two Table

- Chi-square for previous example:

$$\frac{(|(38*17 - 29*7| - 91/2)^2 (91-1))}{(67*24*45*46)} = 4.27 \quad p < 0.05$$

- Used to compare the outcome of two groups
- Commonly used in any treatment vs. control study (ex: drug trials)

33

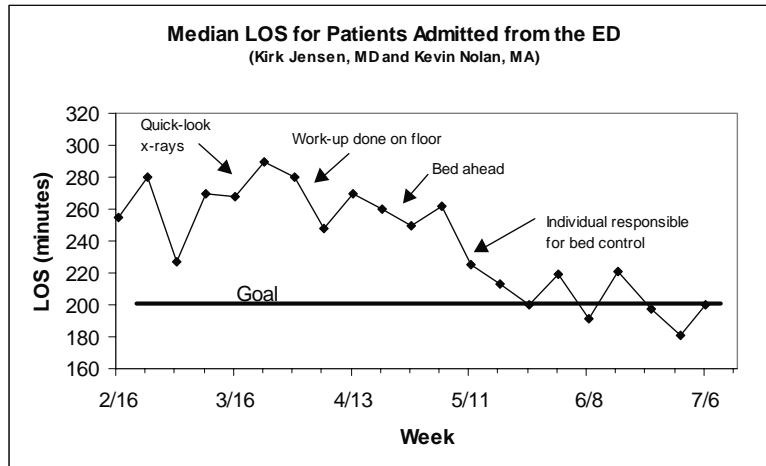


Run Charts and Control Charts

- A run chart of an indicator over time is a powerful, but simple method for studying a process
- A control chart is a run chart with common cause limits added
- Both charts can be used to identify 'common cause' or 'special cause' variation
- May be used to identify affect of an intervention on a process

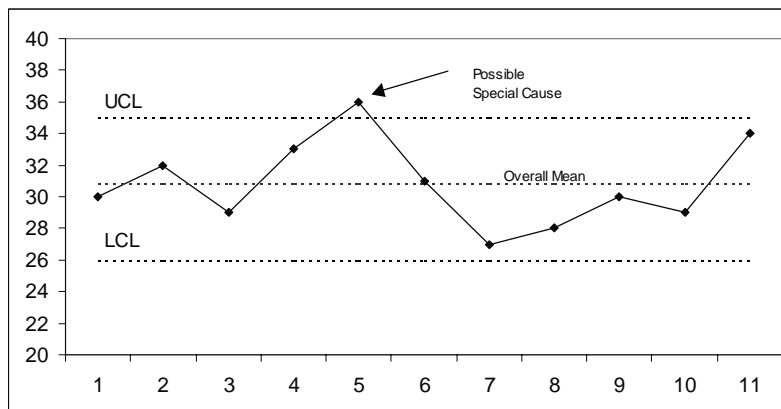
34

Example (Run Chart)



35

Example (Control Chart)



36

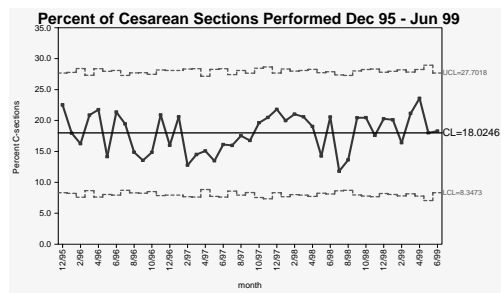
Collecting Data Over Time

- Control limits represent a range around the average of a *STABLE* process.
- Plot the data in a run chart and look for patterns before using control limits. Automatic use of a control chart can complicate a situation.
- Run charts can detect process shifts and other trends that render the summary statistics of a control chart meaningless.

37

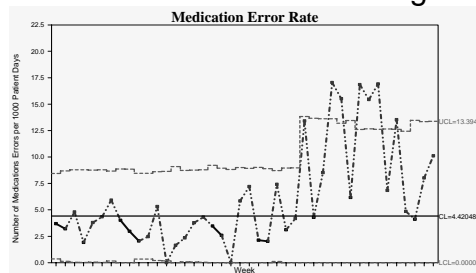
“Common Cause” Variation

- *Stable* and *predictable* pattern of variation
- The result of regular, natural, or ordinary causes
- Affects all the outcomes of a process
- Also known as “random”, “unassignable”, and “controlled” variation



“Special Cause” Variation

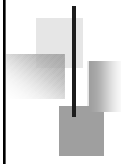
- Unstable and unpredictable pattern of variation
- The result of irregular or unnatural causes that are not inherent in the process
- Affects some, but not necessarily all aspects of the process
- Also known as “non-random” or “assignable” variation



Common Cause does not mean “good” variation.

It only means that the process is stable and predictable. For example, if a patient’s systolic blood pressure averaged around 165 and was usually between 160 and 170 mmHg, this might be stable and predictable but completely unacceptable.

Similarly, **Special Cause** variation should not be viewed as “bad” variation. You could have a special cause that represents a very good result (e.g., a low turnaround time), which you would want to emulate. Special Cause merely means that the process is unstable and unpredictable.

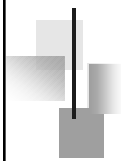


Displaying Data in Charts

- Number of Data Points: For a run or control chart, at least 7 data points (preferably 12 points) are recommended before you can draw meaningful conclusions about the data over time.

- Trends: It is generally acceptable to define a trend as at least seven consecutive data points in either ascending or descending order. The term “trend” is often misused. Avoid using computer-generated “trend” lines, as they are strongly influenced by outlying data points.


41



Collecting Data Over Time

- Control charts and run charts do not answer the following questions:
 - What is the reason for a special cause?
 - Should a common cause process be improved?
 - What should I do to improve the process?
- The answers to these questions will not be found in the charts but in the knowledge of the team!

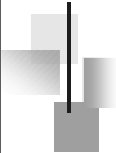
42



Statistical vs. Practical

- Statistical significance is in some measure a product of sample size
- Trivial results may be 'statistically significant' if the sample size is large enough
- Practical significance has less to do with statistics and is a more matter of judgment:
 - Does this make a difference to me?
 - Is it 'significant' to my work?

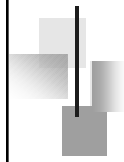
43



Statistical vs. Practical

- Statistical significance does not always imply practical significance.
- For example, survey data may indicate that we score 92% excellence in patient satisfaction and we increase to 94% excellence. If we have a large enough sample size, this may be statistically significant, but the percent increase is rather small in practical terms.

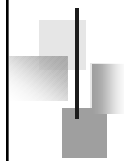
44



Statistical vs. Practical

- Even though some data collected for some quality improvement projects may not need minimum criteria for statistical significance, there is a lot we can learn from it of practical significance.
- We may still learn from the abstracting process or from investigating the clinical process being measured, even if there are no statistically sound measurements.

45



Summary statistics are not always a good choice...

- Summary statistics, presented in tabular format, are not helpful when assessing the impact of process improvement.
- Means can be influenced by a few, extreme observations.
- Variation in your data needs to be appreciated visually.

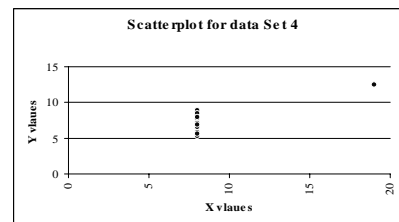
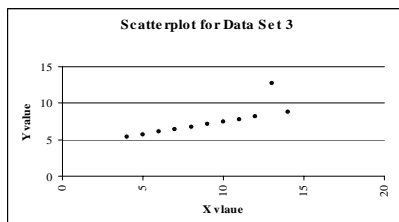
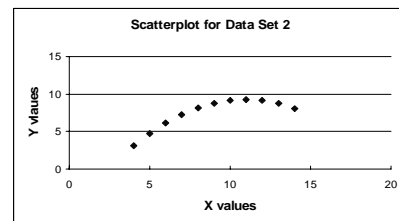
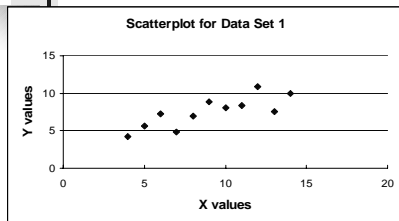
46

Consider an example...

Data Set 1		Data Set 2		Data Set 3		Data Set 4	
X	Y	X	Y	X	Y	X	Y
10.00	8.04	10.00	9.14	10.00	7.46	8.00	6.58
8.00	6.95	8.00	8.14	8.00	6.77	8.00	5.76
13.00	7.58	13.00	8.74	13.00	12.74	8.00	7.71
9.00	8.81	9.00	8.77	9.00	7.11	8.00	8.84
11.00	8.33	11.00	9.26	11.00	7.81	8.00	8.47
14.00	9.96	14.00	8.10	14.00	8.84	8.00	7.04
6.00	7.24	6.00	6.13	6.00	6.08	8.00	5.25
4.00	4.26	4.00	3.10	4.00	5.39	19.00	12.50
12.00	10.84	12.00	9.13	12.00	8.15	8.00	5.56
7.00	4.82	7.00	7.26	7.00	6.42	8.00	7.91
5.00	5.68	5.00	4.74	5.00	5.73	8.00	6.89
Average	Average	Average	Average	Average	Average	Average	Average
9.0	7.5	9.0	7.5	9.0	7.5	9.0	7.5

All four of these datasets have the same N,
and the same mean for X and Y, *but...*

...look at their respective scatter plots.



Now what do you conclude??



Free Statistical Software

- There are a number of free, simple statistical software programs available in addition to the programs that you can purchase, such as SAS, SPSS, or STATA.
- One example of a website with free statistical programs available to download:
<http://statpages.org/javasta2.html>
(I have used the Smith Statistical Package (SSP) available on this site)

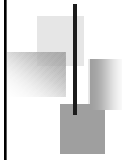
49



***In God we trust.
All others must use data.***

- W. Edward Deming

50



Trista Johnson Ph.D., MPH
Director of Performance Measurement
and Analysis Allina Hospitals & Clinics
Phone (612) 262-6015
Trista.Johnson@allina.com