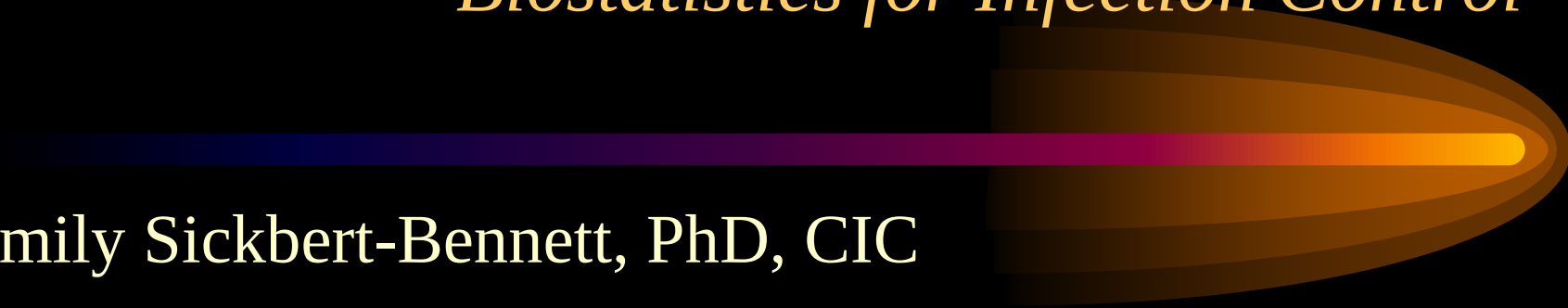


Making the Most of Your Surveillance Data: Biostatistics for Infection Control



Emily Sickbert-Bennett, PhD, CIC

Associate Director of Hospital Epidemiology, UNC Health Care
Research Assistant Professor of Medicine-Infectious Diseases,
UNC School of Medicine

Adjunct Assistant Professor of Epidemiology,
Gillings School of Global Public Health

Statistics

The margin of
error...

17 in every 100 people...

Men are at 3
times higher
risk...

Numbers that describe
the health of the
population

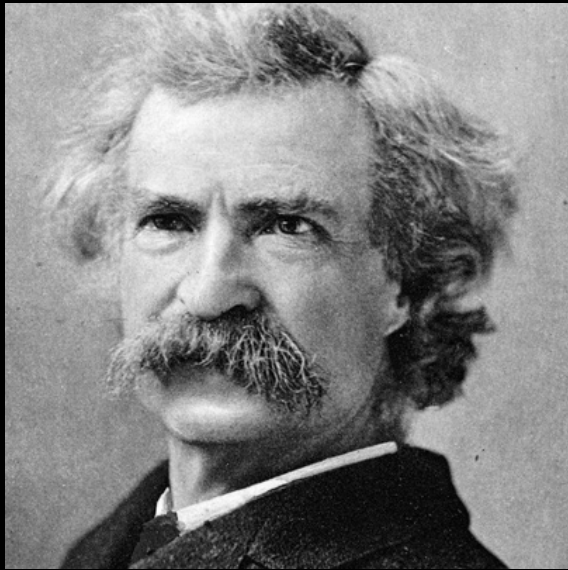
1 in 9 children...

39% OF THE
POPULATION...

The science used to
interpret these numbers.

Risk of dying is 8 times higher
among...

There is a
statistically
significant
difference...



**“There are 3 kinds of lies.
Lies, damned lies, and
statistics.”**

~Popularized by Mark Twain

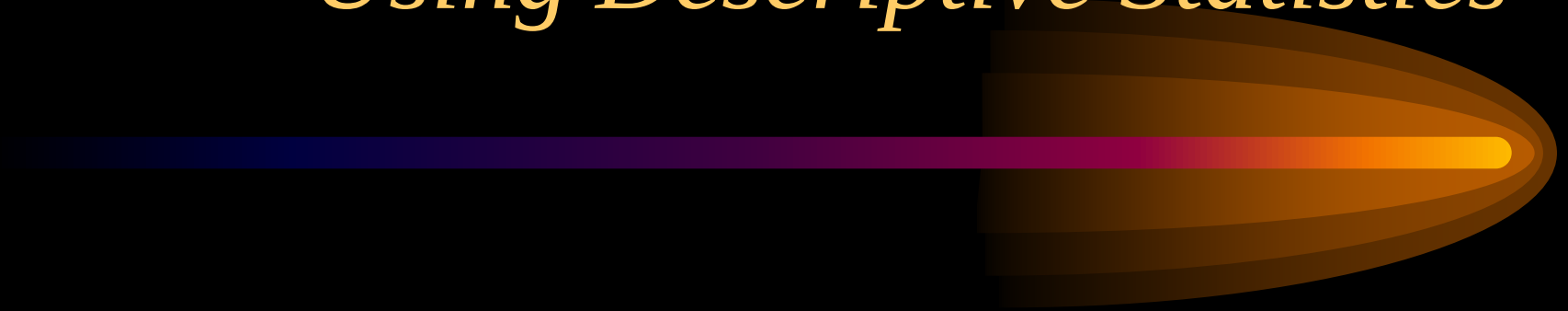
- Describes the persuasive power of numbers, particularly the use of statistics, to bolster weak arguments, and the tendency of people to disparage statistics that do not support their positions.

Outline

A decorative graphic consisting of a horizontal bar with a color gradient from dark blue on the left to bright yellow on the right. To the right of the bar is a large, stylized arrow pointing to the right, filled with a brown-to-yellow gradient.

- Describe Surveillance Data
- Display and Interpret Surveillance Data
- Determine the Significance of Changes to Surveillance Data

Describing Surveillance Data Using Descriptive Statistics



Absolute Measures

- Simplest type of measurement
- Also known as counts or frequencies
- Example:
 - Hospital A: 25 patients with norovirus
 - Hospital B: 10 patients with norovirus
- Is norovirus worse at Hospital A?

Relative Measures

- Includes a denominator
- Useful for comparisons
- Examples:
 - 16 cases of *C. difficile* out of 1000 patients
 - 1 positive *C. difficile* test out of 7 samples tested

Absolute versus Relative

Example: Norovirus activity at different hospitals

- Absolute measures
 - Hospital A: 25 patients ill
 - Hospital B: 10 patients ill
- Relative measures
 - Hospital A: 25 ill per 1000 patients = 0.025 or 2.5%
 - Hospital B: 10 ill per 250 patients = 0.040 or 4%

Descriptive Statistics



- Measures of Rates and Ratios
 - *Rate: How fast disease occurs in a population.*
 - *Ratio: How much disease compared to standard.*
- Measures of Central Tendency
 - *Central Tendency: How well the data clusters around an average value.*
- Measures of Dispersion
 - *Dispersion: How widely your data is spread from the average.*

What Makes a Rate?

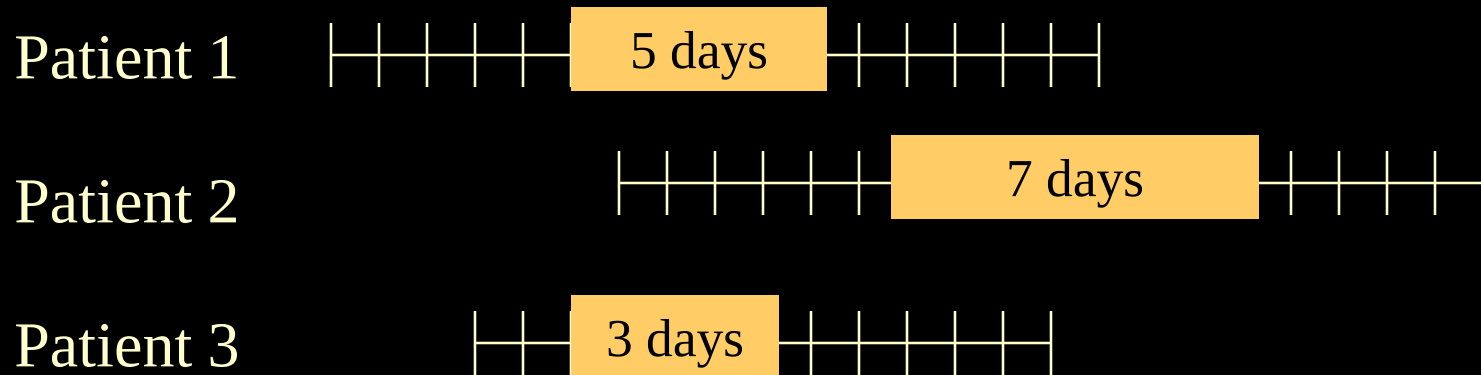
1. Numerator (top number)
 - *e.g., number of infections*
2. Denominator (bottom number)
 - *e.g., number of patients [proportion]*
 - *e.g., number of patient-days, number of device-days [incidence density/rate]*
3. Time Frame
 - *e.g., day, week, month*

Denominators



- Represent the population at risk of becoming part of the numerator
- Often, the most difficult data to obtain, but essential for comparisons
- Ideally, should incorporate time and can account for risk factors such as device use (e.g., device-days), length of stay (e.g., patient-days)

What is a Patient/Device-Day?



=15 patient-days, device-days, etc.

- Gives more information than simply—3 patients
- Strategies: e.g., count how many at 9 am

Rate Measures



- Prevalence
- Incidence
- Attack Rate

Prevalence

- Prevalence: the total number of cases of disease existing in a population at a point in time.
 - *e.g., # of MRSA cases per population on March 8*

$$\frac{\text{Count of existing cases}}{\text{Number of people at risk}} \times \text{constant (e.g., 100 or 1000)} =$$

Incidence

- Incidence: the number of new cases of disease in a population over a period of time.
 - *e.g., # of new MRSA cases per population during March*

$$\frac{\text{Count of new cases}}{\text{Number of people at risk}} \times \text{constant (e.g., 100 or 1000)} =$$

Attack Rate

- Attack Rate: the number of new cases of disease out of the population at risk.
 - Related to incidence but always uses 100 as the constant, so it is expressed as a percent.
 - Often used for outbreaks or clusters that occur over a short period of time
 - *e.g., % of patients with MRSA during outbreak in Med ICU in March*

$$\frac{\text{Count of new cases}}{\text{Number of people at risk}} \times 100 =$$

Example 1:

- You perform HAI surveillance for ventilator associated pneumonias (VAP) and central line associated blood stream infections (BSI) in your 12 bed intensive care unit.
- In March, you identify 2 new VAPs, 4 new CLBSIs and 3 new respiratory infections (not ventilator associated).

Example 1:

- The admitting department tells you that in March there were 89 patients in the unit with 311 patient-days.
- Respiratory care tells you that they provided 162 ventilator-days of care to 47 patients in March.
- You count the central line-days and find 284 line-days in 84 patients in March.

Example 1:

- In March, what was the VAP rate?
 - Incidence or prevalence?
 - Numerator?
 - Denominator?
 - Units?

Example 1:

- In March, what was the CLA-BSI rate?
 - Incidence or prevalence?
 - Numerator?
 - Denominator?
 - Units?

Example 1:

- In March, what was overall infection rate?
 - Incidence or prevalence?
 - Numerator?
 - Denominator?
 - Units?

Example 1:



- On April 7, you were worried about the BSI rate so you return to the unit to do a “spot check” on all of the patients for a BSI.
- At that time with a census of 12, you reviewed 11 charts and found 1 healthcare associated BSI.

Example 1:

- On April 7th, what was the BSI infection rate at the time of your spot check?
 - Incidence or prevalence?
 - Numerator?
 - Denominator?
 - Units?

What Makes a Standardized Infection Ratio (SIR)?

1. Numerator (top number)
=number of observed infections
2. Denominator (bottom number)
=number of expected or predicted infections
 - Number of predicted infections =
calculated based on your hospital's number of
procedures, device days, risk factors, nursing units
*compared to a standard infection rate (e.g.,
historical data, state data, national data)*

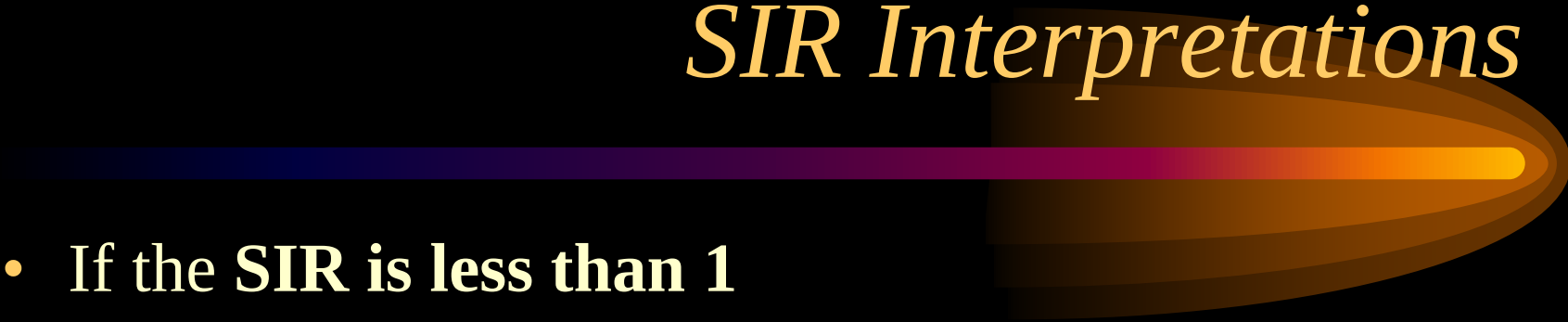
Standardized Infection Ratio

- $SIR = \frac{\text{\# observed infections}}{\text{\# predicted infections}}$
- $SIR > 1.0 \rightarrow \text{more infections than predicted}$
- $SIR < 1.0 \rightarrow \text{fewer infections than predicted}$
- *~LOWER SIRs are BETTER~*

SIR Interpretations

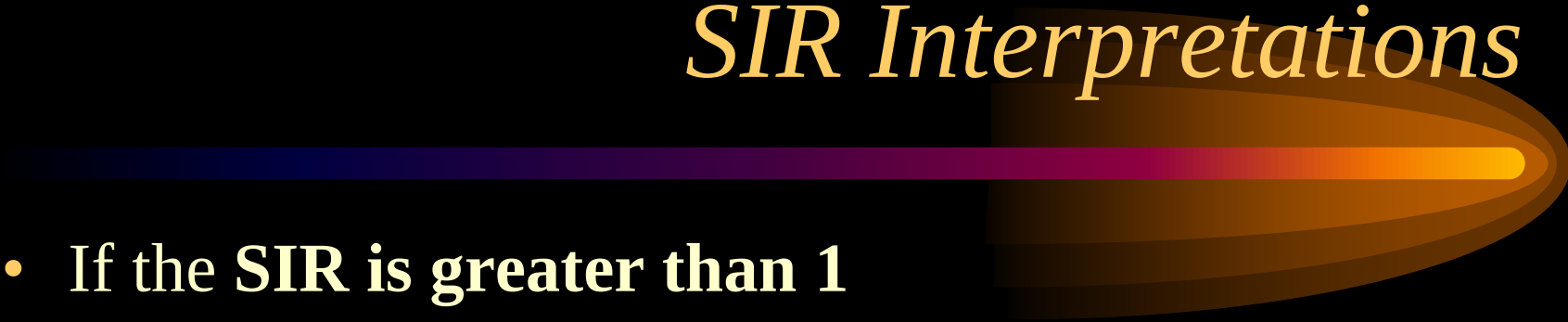
- **SIR=1**
- The number of infections is the same as the number of expected infections
- No progress has been made in reducing infections since the baseline period or compared to another standard population (e.g., all NC, all US).

SIR Interpretations



- If the **SIR is less than 1**
 - Fewer infections than predicted based on standard or baseline data
 - Infection reduction/prevention compared to standard or baseline data
 - 1 minus the SIR = percent reduction:
For example, a SIR of 0.80 means that there was a 20 percent reduction from the standard population or baseline time period

SIR Interpretations



- If the **SIR is greater than 1**
 - More infections than predicted based on standard or baseline data
 - Infections are increased compared to standard or baseline data
 - SIR minus 1 = percent increase:
For example, a SIR of 1.25 means that there was a 25 percent increase from the standard population or baseline time period

Example 1: SIR

- CLABSI rate = 4 CLABSI/284 line days
- Predicted Infections = 0.57
- What is the SIR?
- How would you explain the SIR to your administrator?

2017: New SIRs and Rebaseline



- More recent data (2015 as baseline year) so increase in most SIRs is expected
- Benefits:
 - Mucosal Barrier Injury for CLABSI are excluded
 - Present at Time of Surgery for SSI are excluded

Descriptive Statistics



- Measures of Rates
 - *Rate: How fast disease occurs in a population.*
 - *Ratio: How much disease compared to standard.*
- Measures of Central Tendency
 - *Central Tendency: How well the data clusters around an average value.*
- Measures of Dispersion
 - *Dispersion: How widely your data is spread from the average.*

Measures of Central Tendency

- **Mean:** average of a group of numbers
- **Median:** middle number in an ordered group of numbers
- **Mode:** most common value in a group of numbers

Hey diddle diddle,
the median's the middle;

YOU ADD AND DIVIDE FOR THE MEAN.

The mode is the one that appears the most,
and the range is the difference between.

Descriptive Statistics

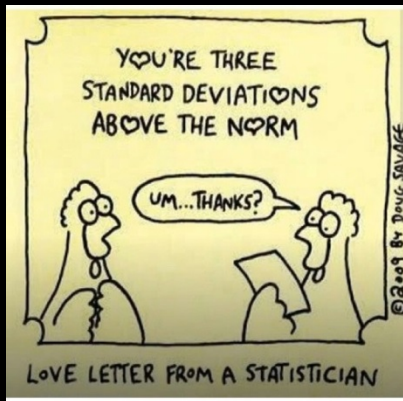


- Measures of Rates
 - *Rate: How fast disease occurs in a population.*
 - *Ratio: How much disease compared to standard.*
- Measures of Central Tendency
 - *Central Tendency: How well the data clusters around an average value.*
- Measures of Dispersion
 - *Dispersion: How widely your data is spread from the average.*

Measures of Dispersion

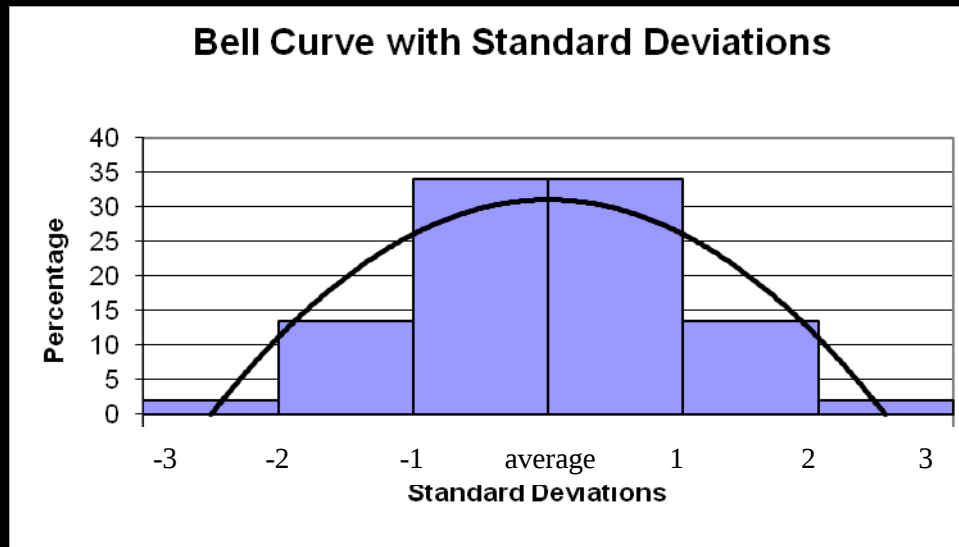


- Range: the largest value minus the smallest value
- Standard deviation: describes the variability or dispersion in the data set



Standard Deviation

- In an normally distributed data set,



68% of values ± 1 SD

95% of values ± 2 SD

99% of values ± 3 SD

Example 2:



- Your administrator is becoming concerned about the impact of healthcare associated infections on the length of stay in your ICU.
- She has asked you to provide her with some data to confirm her suspicions.

Example 2:

- Over the last 3 months you have identified a series of 31 ventilator-associated pneumonias with the total length of stay for each ICU patient as follows:

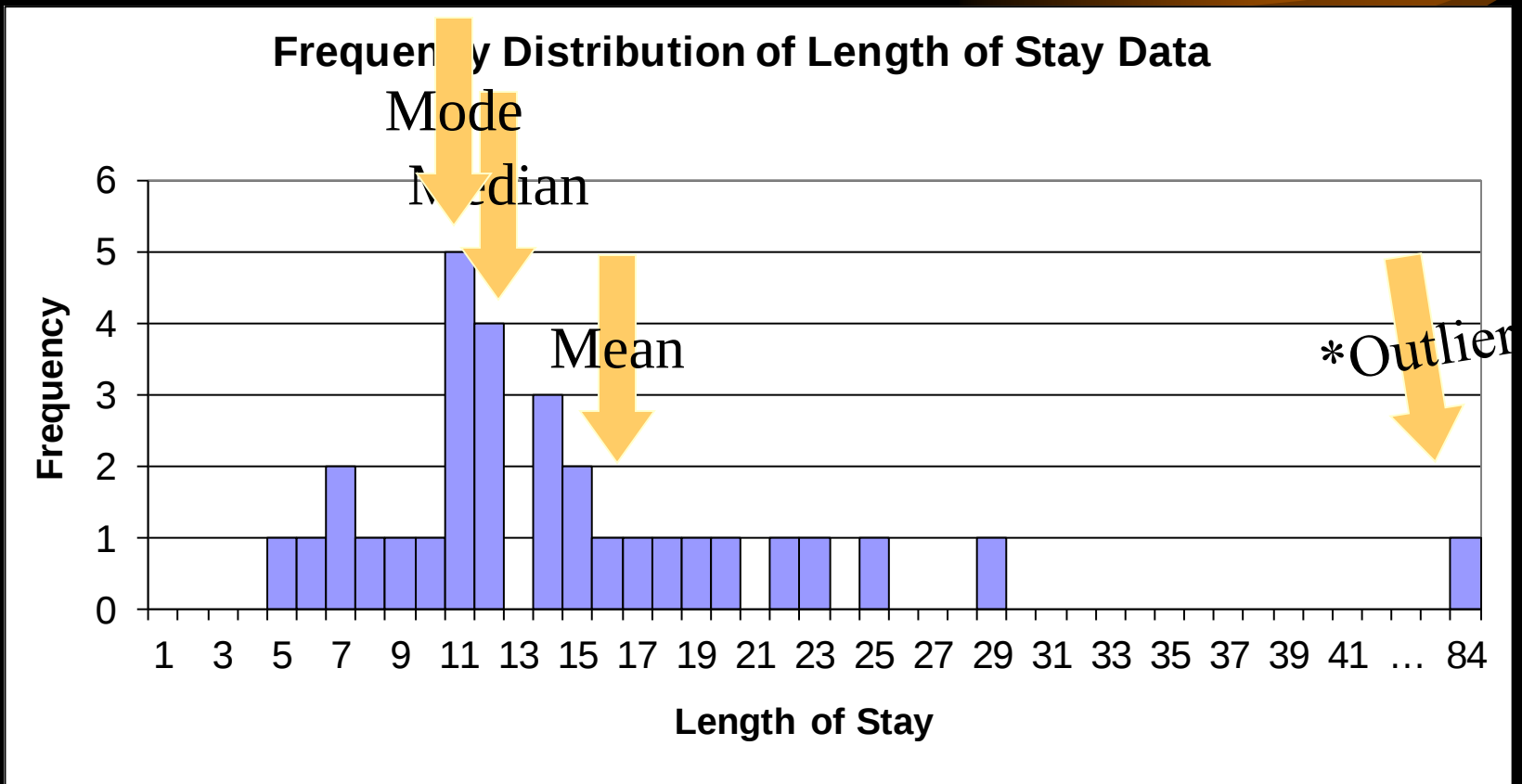
9, 7, 14, 11, 12, 22, 15, 10, 29, 16, 11, 7, 5,
12, 17, 25, 14, 14, 15, 23, 20, 11, 12, 18, 19,
11, 8, 6, 84, 12, 11

Example 2:

- What is the:
 - Mean?
 - Median?
 - Mode?
 - Range?

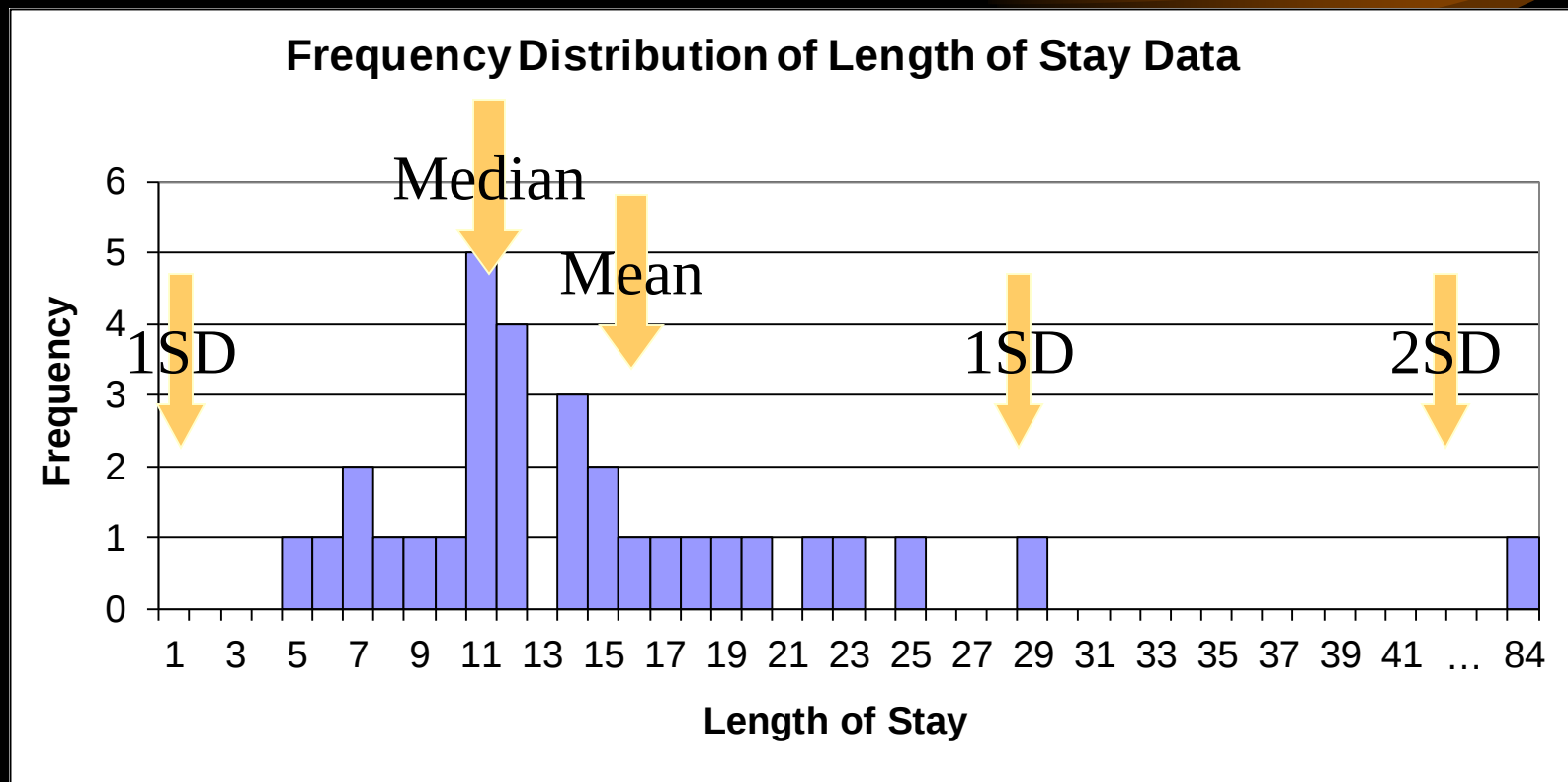
HINT: 5, 6, 7, 7, 8, 9, 10, 11, 11, 11, 11, 11, 12, 12, 12, 12,
14, 14, 14, 15, 15, 16, 17, 18, 19, 20, 22, 23, 25, 29, 84

Example 2: Central Tendency



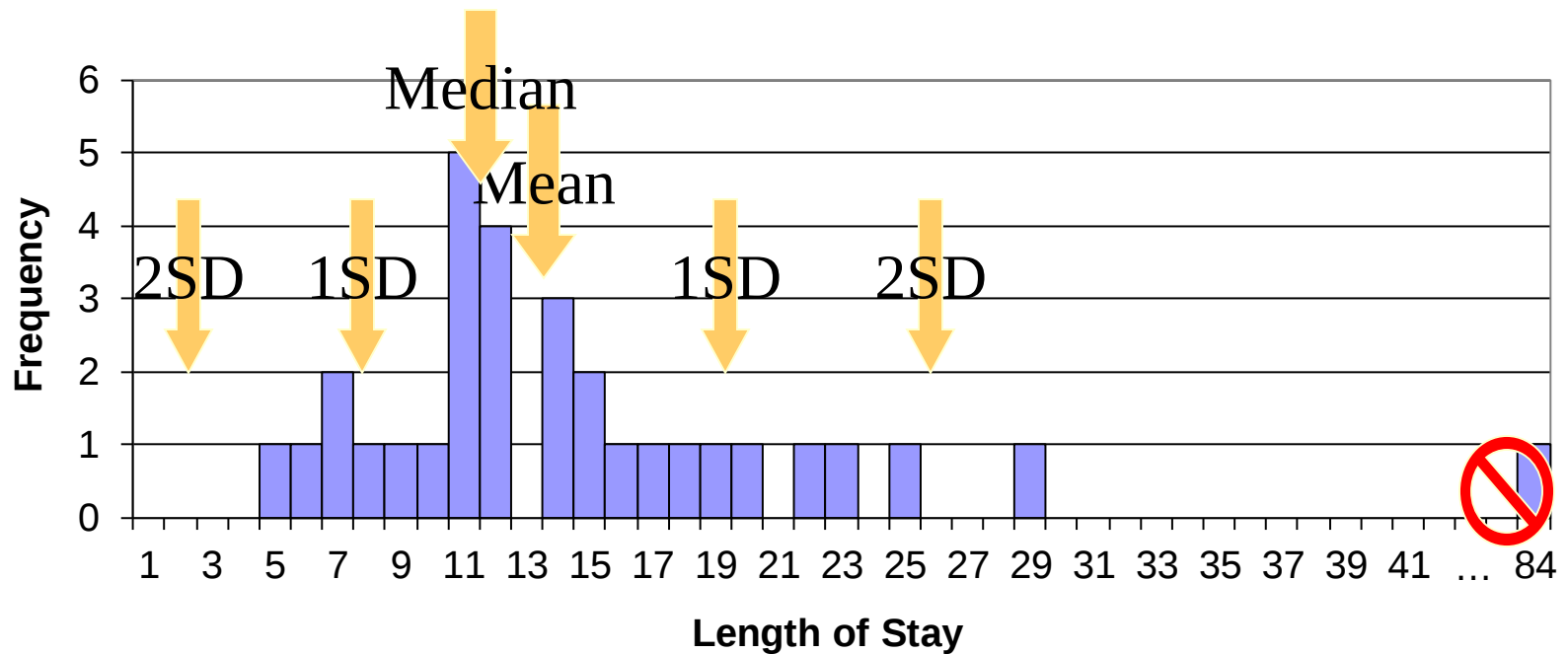
*outlier: a value that falls outside the overall pattern.

Example 2: Dispersion

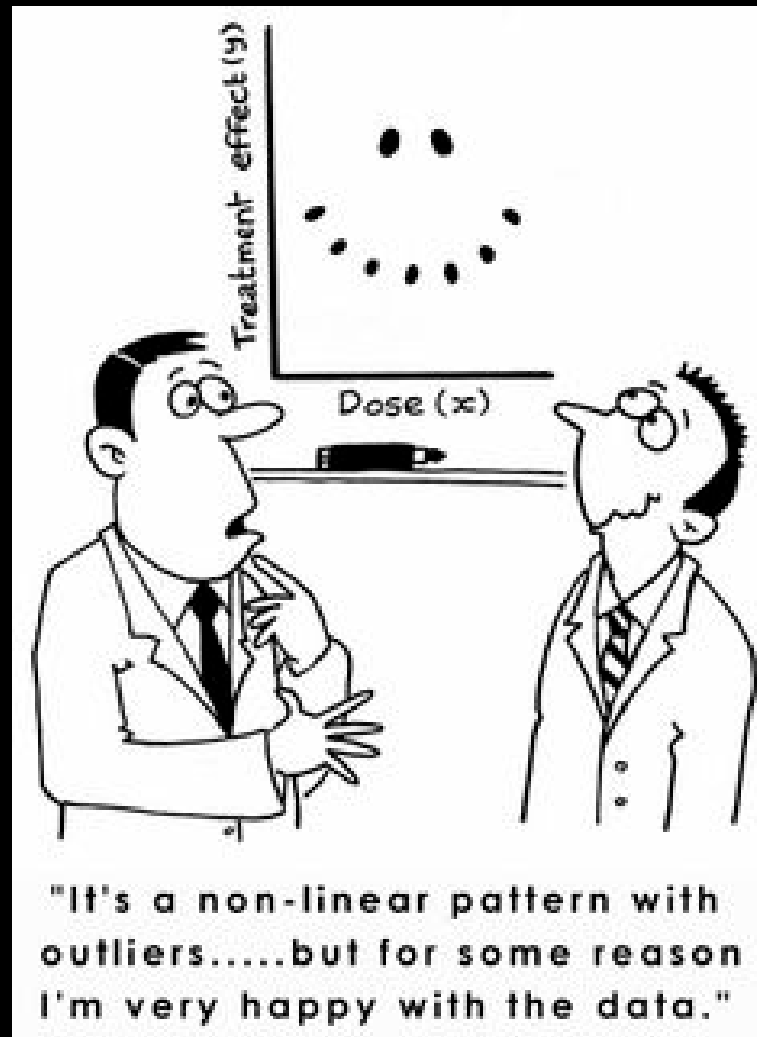


Example 2: Dispersion

Frequency Distribution of Length of Stay Data



Displaying Surveillance Data



Displaying and Interpreting Surveillance Data



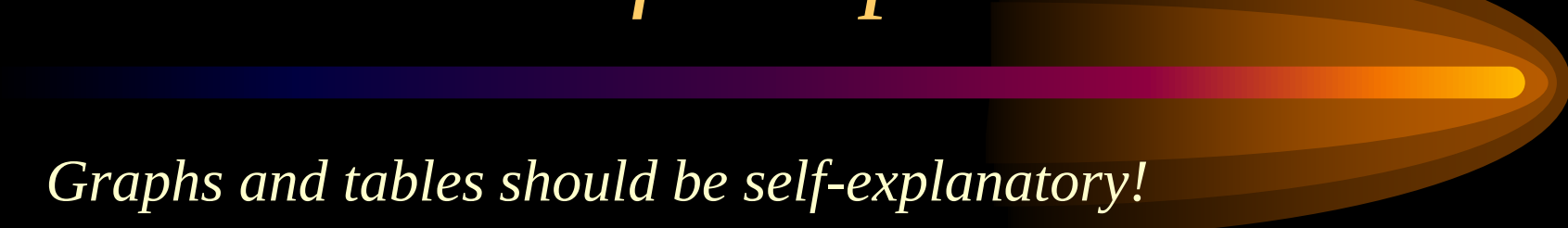
- Graphs: a visual representation of data on a coordinate system (e.g., two axes)
- Tables: a set of data arranged in rows and columns

Data Types



- Quantitative variables: numerical values
 - *(e.g., number of infections, number of patients)*
- Categorical variables: descriptive groups or categories
 - *(e.g., units in the hospitals, occupational groups)*

Features of Graphs and Tables



Graphs and tables should be self-explanatory!

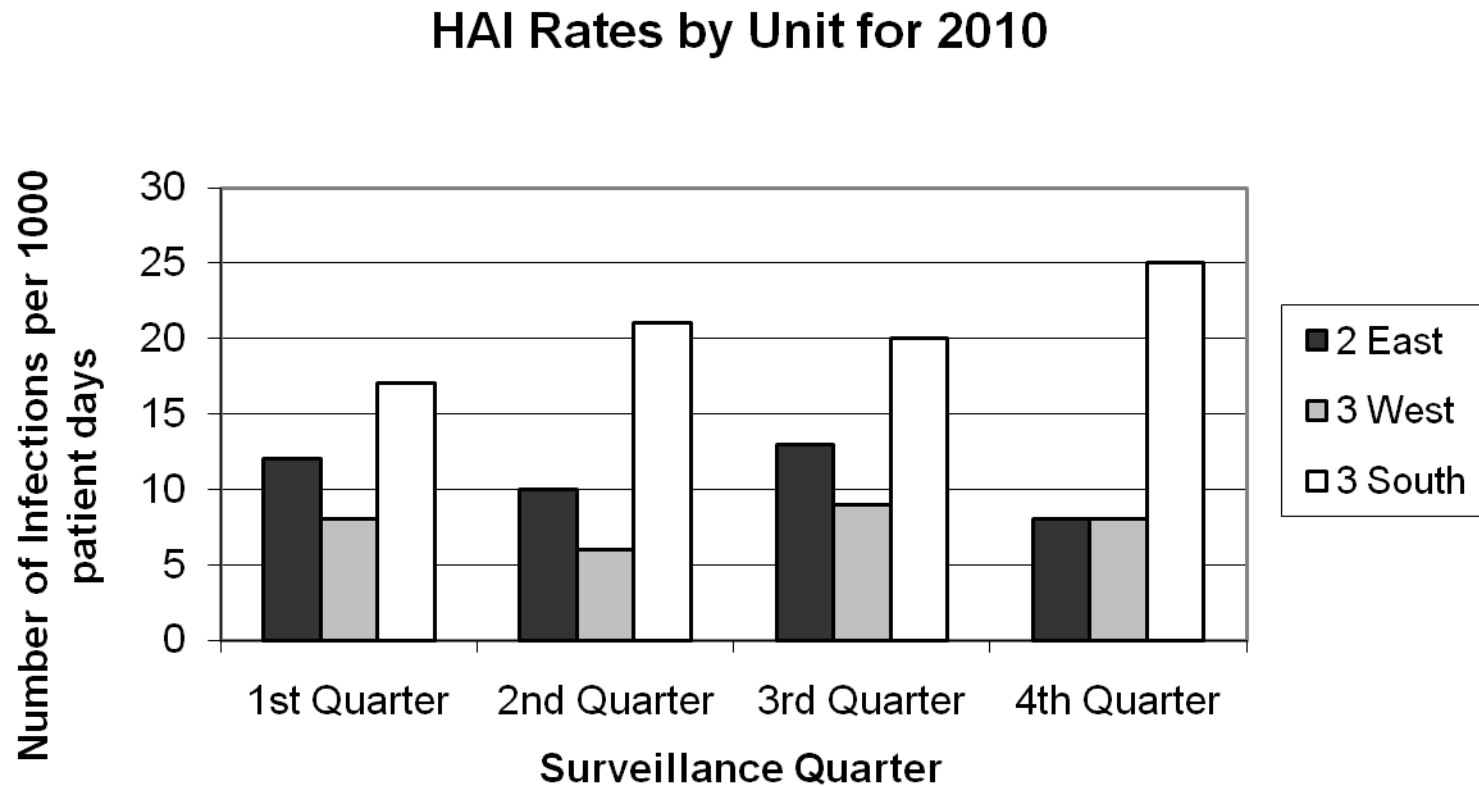
- Clear, concise title: describes person, place, time
- Informative labels: axes, rows, columns
- Appropriate intervals for axes
- Coded and labeled legends or keys
- Use footnotes to:
 - Explain codes, abbreviations, and symbols
 - Note exclusions
 - Note data source

Graph Types

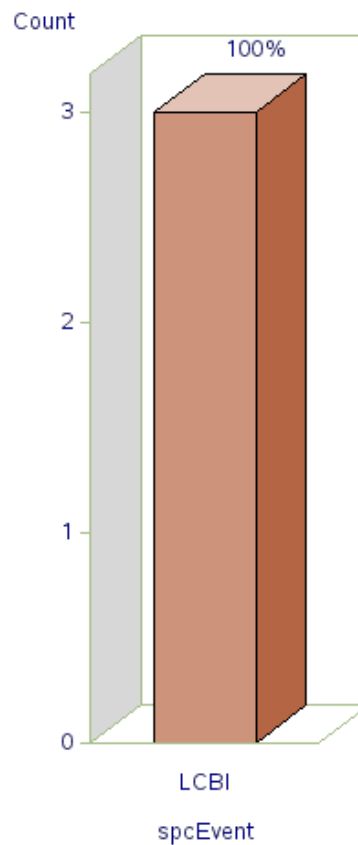


- Bar Graphs
 - *E.g, Histograms (shown in previous example)*
 - *E.g., Comparison between categories*
 - *E.g., Epidemic Curves*
- Line Graphs
 - *E.g., To show trends over time*
- Pie Charts
 - *E.g., As a percentage of a whole*

Bar Graph



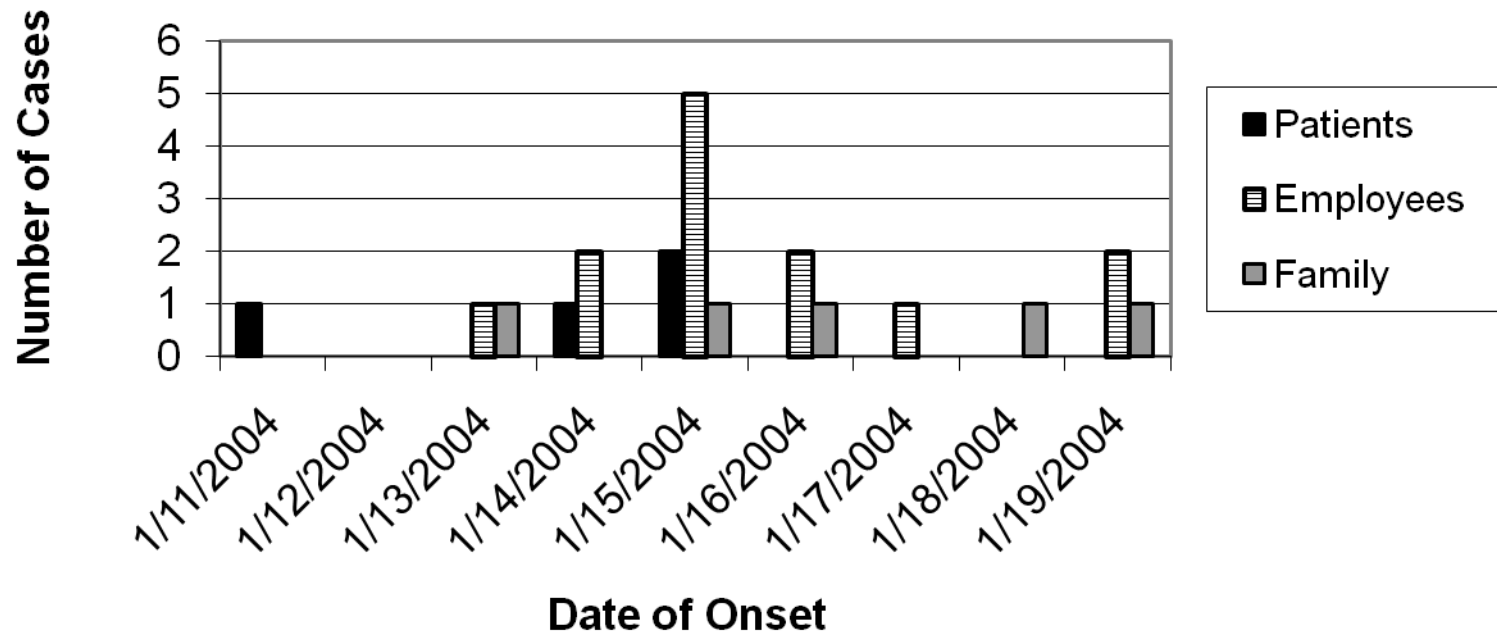
National Healthcare Safety Network
 Bar Chart for All Central Line-Associated BSI Events
 As of: March 29, 2013 at 5:06 PM
 Date Range: All CLAB_EVENTS
 location=SICU



Data contained in this report were last generated on March 22, 2013 at 3:40 PM.

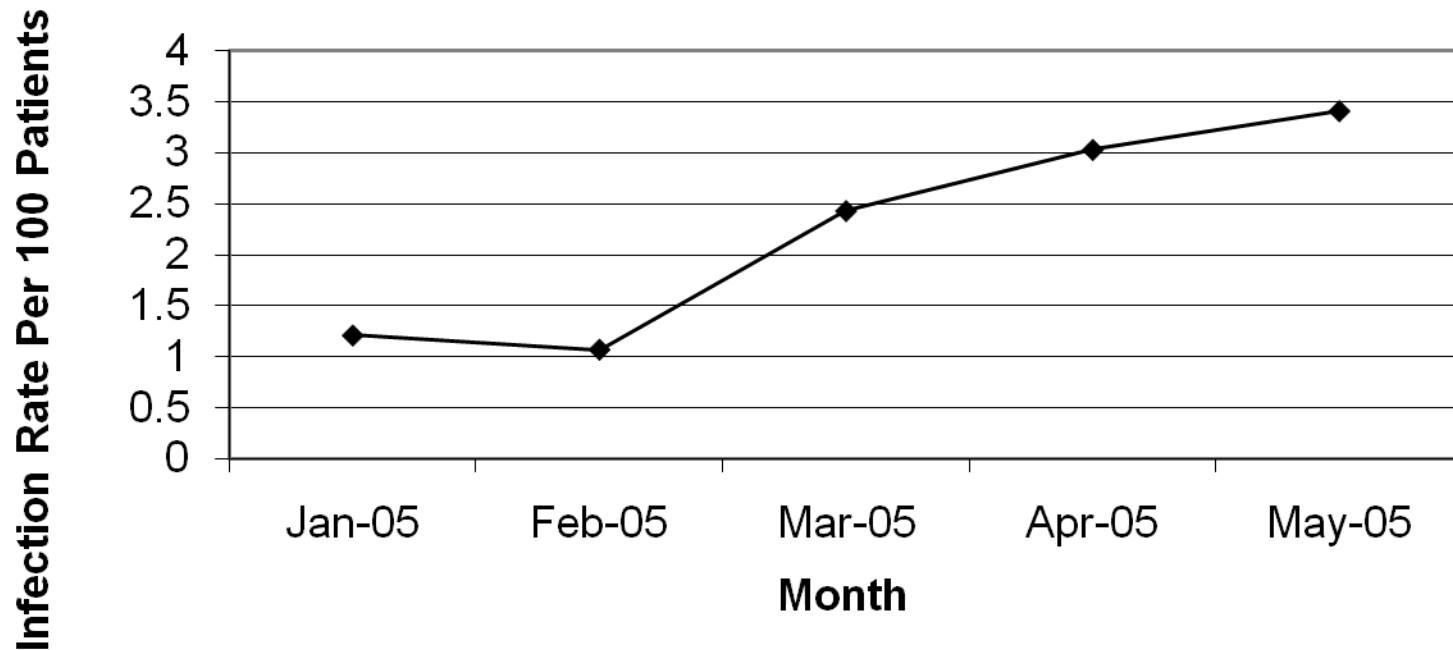
Epi Curve

Epidemic Curve for Gastroenteritis Outbreak on 5th Floor, January 2004



Line Graph

**Surgical Site Infections in Hospital X
January 2005- May 2005**



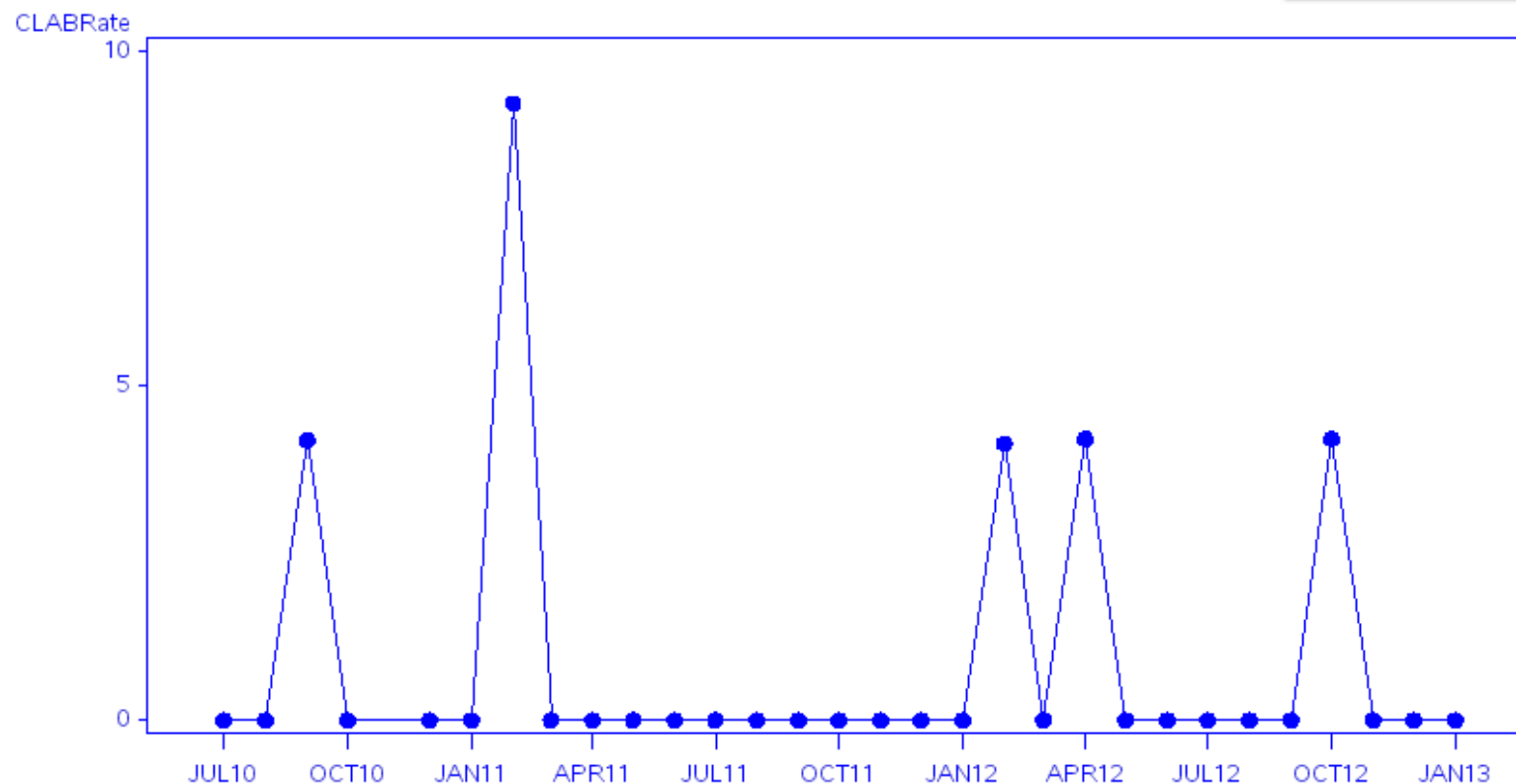
Data contained in this report were last generated on March 22, 2013 at 3:40 PM.

National Healthcare Safety Network

Run Chart for Central Line-Associated BSI Data for ICU-Other
As of: March 22, 2013 at 4:15 PM Date Range: All CLAB_RATESICU
orgID=16195 location=CTICU

●●● CLA BSI Rate

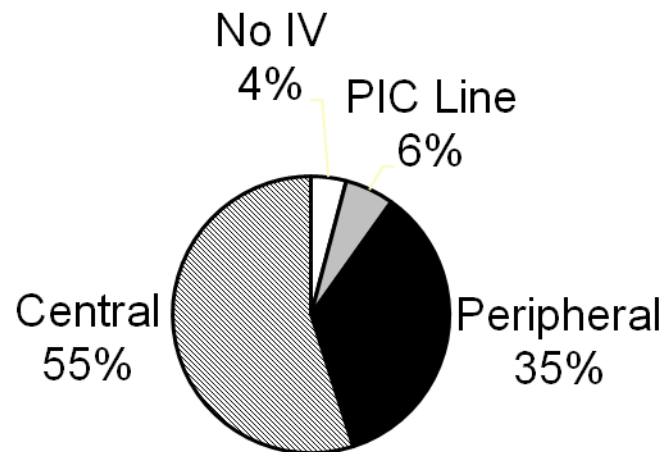
Plot of CLABRate by summaryym



NOTE: Strata with fewer than 2 data points have been omitted.

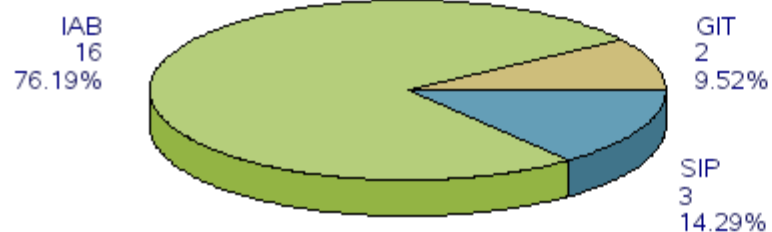
Pie Chart

Distribution of Primary Bloodstream Infections by Device Type at Hospital X for 2009

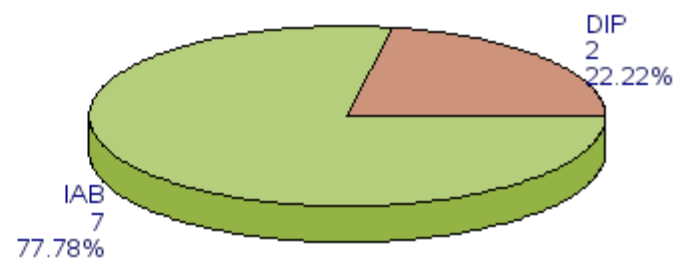


National Healthcare Safety Network
 Pie Chart for All Surgical Site Infection Events
 As of: March 22, 2013 at 4:21 PM
 Date Range: All SSI_EVENTS
 FREQUENCY of spcEvent

procCode=COLO



procCode=HYST



Pie chart of spcEvent

Tables

Number of Newly Diagnosed Cases by Age, United States, 2010

Age Group (Years)	Number of Cases
0-4	1242
5-14	1081
15-24	2482
25-44	8153
45-64	10916
65+	7124
Total	30998

National Healthcare Safety Network**Rate Table for Central Line-Associated BSI Data for ICU-Other**

As of: March 22, 2013 at 4:17 PM

Date Range: All CLAB_RATESICU

orgID=16195 loccdc=IN:ACUTE:CC:C

location	summaryYM	CLABCount	numCLDays	CLABRate	CLAB_Mean	IDR_pval	IDR_pctl	numPatDays	LineDU	LineDU_Mean	P_pval	P_pctl
CICU	2010M07	0	223	0.000	1.1	0.7804	25	268	0.832	0.42	0.0000	94
CICU	2010M08	0	290	0.000	1.1	0.7243	25	294	0.986	0.42	0.0000	98
CICU	2010M09	1	236	4.237	1.1	0.2308	96	262	0.901	0.42	0.0000	95
CICU	2010M10	0	276	0.000	1.1	0.7357	25	328	0.841	0.42	0.0000	94
CICU	2010M12	0	253	0.000	1.1	0.7548	25	269	0.941	0.42	0.0000	96
CICU	2011M01	1	282	3.546	1.1	0.2692	93	304	0.928	0.42	0.0000	96
CICU	2011M02	0	298	0.000	1.1	0.7179	25	314	0.949	0.42	0.0000	97
CICU	2011M03	0	241	0.000	1.1	0.7649	25	274	0.880	0.42	0.0000	95
CICU	2011M04	1	238	4.202	1.1	0.2325	95	272	0.875	0.42	0.0000	95
CICU	2011M05	0	213	0.000	1.1	0.7891	25	281	0.758	0.42	0.0000	92
CICU	2011M06	0	237	0.000	1.1	0.7683	25	253	0.937	0.42	0.0000	96
CICU	2011M07	0	161	0.000	1.1	0.8361	25	227	0.709	0.42	0.0000	91
CICU	2011M08	0	218	0.000	1.1	0.7847	25	280	0.779	0.42	0.0000	92
CICU	2011M09	0	195	0.000	1.1	0.8051	25	295	0.661	0.42	0.0000	88
CICU	2011M10	0	239	0.000	1.1	0.7666	25	316	0.756	0.42	0.0000	92
CICU	2011M11	1	230	4.348	1.1	0.2257	96	287	0.801	0.42	0.0000	93
CICU	2011M12	0	228	0.000	1.1	0.7760	25	317	0.719	0.42	0.0000	91

Interpreting Surveillance Data



NHSN data summary, 2013

Urinary catheter associated UTI Rate*					Percentile				
Types of Location: Critical care units	No. Loca- tions	No. of CAUTI	Urinary catheter days	Pooled Mean	10%	25%	50% (median)	75%	90%
Medical cardiac	384	1494	658,345	2.3	0.0	0.7	1.9	3.4	4.9
Medical/surgical ≤15 beds	1645	2429	1,910,118	1.3	0.0	0.0	0.4	1.7	3.1
Surgical cardiothoracic	453	1715	942,852	1.8	0.0	0.7	1.5	2.4	3.4

* $\frac{\text{Number of CA UTIs}}{\text{Number of urinary catheter days}} \times 1000$

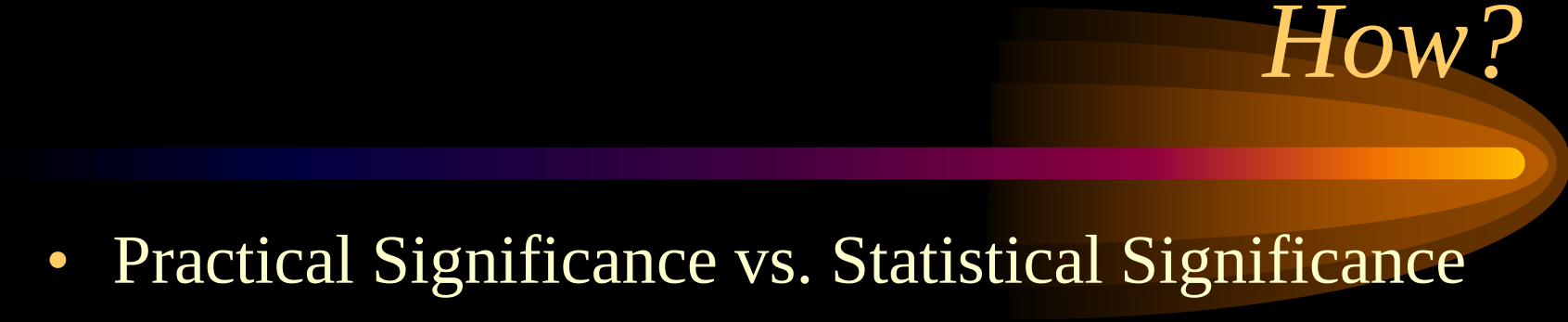
What does this NHSN data summary tell you?

- What is the mean UTI rate in the medical cardiac critical care unit?
 - 2.3 UTIs per 1000 urinary catheter days
- If your medical/surgical (≤ 15 beds) critical care unit has a rate of 1.6 UTIs per 1000 urinary catheter days—between what percentiles is it compared to the NHSN data?
 - Between the 50th-75th percentiles
- If your surgical cardiothoracic critical care unit has a rate of 4.2 UTIs per 1000 urinary catheter days—between what percentiles is it compared to the NHSN data?
 - Greater than the 90th percentile

Determine the Significance of Changes to Surveillance Data



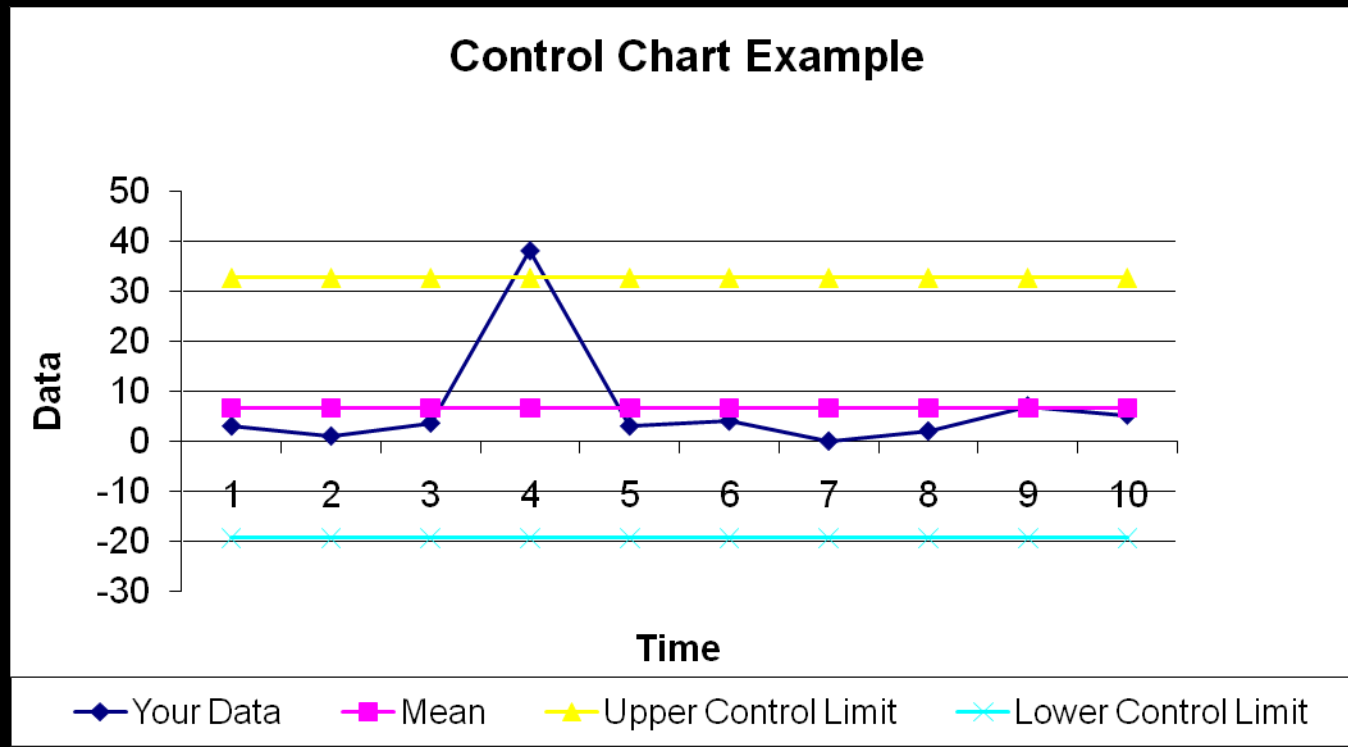
Determine the Significance- How?



- Practical Significance vs. Statistical Significance
- Make comparisons
 - *For example: over time, to other units, to other hospitals (NHSN data)*
 - Remember to choose appropriate data for comparison (*i.e., same denominator units*)
- Apply a type of statistical test
 - *e.g., control charts (for time trends)*
- Other statistical tests and measures
 - P-values
 - 95% confidence intervals

Control Charts

- Tool to determine when infection rates are out of range. *How high is TOO high?*



Control Chart

Example 3:

Month	2015 BSI Rate	Moving Range
1	4.5	--
2	3.2	1.3
3	3.6	0.4
4	3.5	
5	3.0	
6	4.0	
7	4.1	
8	4.6	
9	4.8	
10	5.2	
11	5.7	
12	6.5	

- Find the mean of the BSI rates for the last year.
- Calculate the moving ranges (subtract month 1 from 2, month 2 from 3...) and take absolute values (no negative values).
- Calculate the mean of the moving ranges.

Control Chart

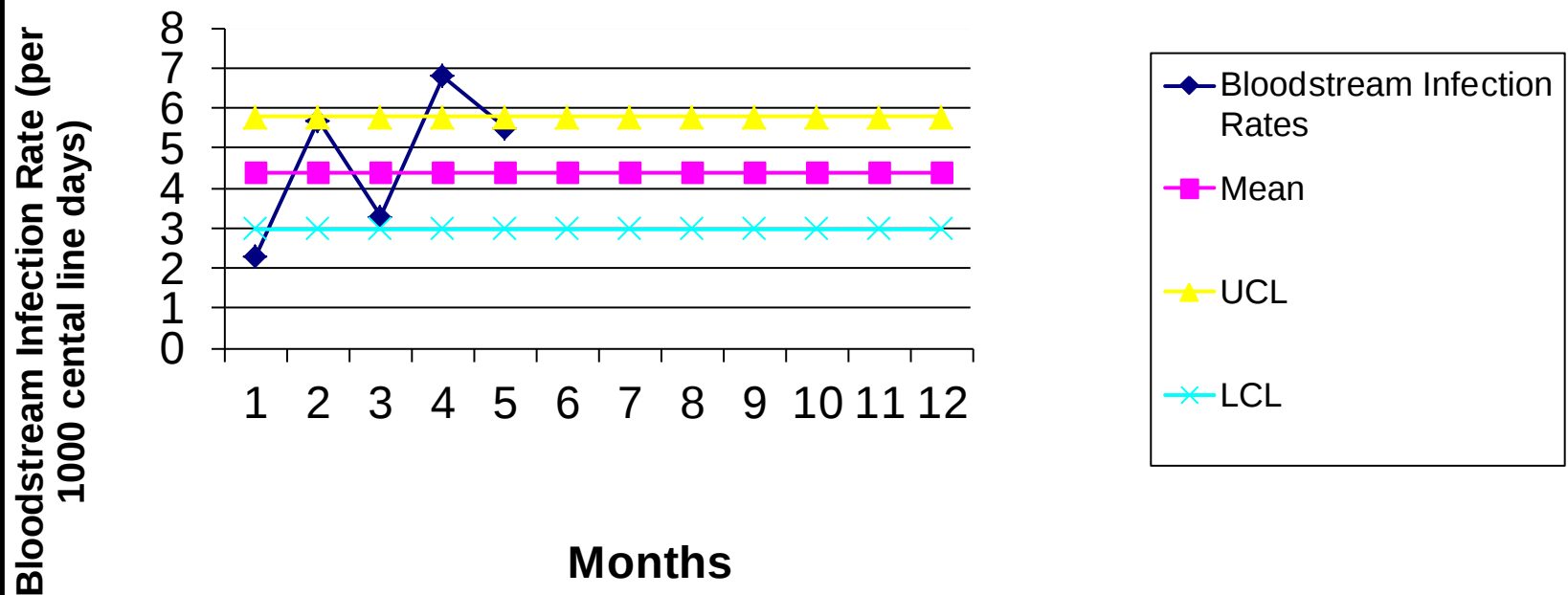
Example 3:

- Calculate Upper Control limit= $\text{Mean} + (2.66 \times \text{Mean of Moving Range})$
- Calculate Lower Control limit= $\text{Mean} - (2.66 \times \text{Mean of Moving Range})$
- Draw horizontal lines at the mean, UCL and LCL based on your historical data
- Then graph your current data and use the limits to identify potential problems.

Control Chart

Example 3:

**Bloodstream Infection Rates
at Hospital for Year 2016**



P-Value

Example:

- “Our study showed that people who washed their hands were less likely to get sick ($P=0.06$) and more likely to be nurses ($P=0.01$).”

Testing Hypotheses and P-Values

- Null hypothesis: values are equal
- Alternative hypothesis: values differ
- These statements are mutually exclusive
 - They cover all possible outcomes
 - In the end, only one can be selected

p-value: *If the null hypothesis is true*, the probability you would find a difference as extreme or more extreme than what you've observed.

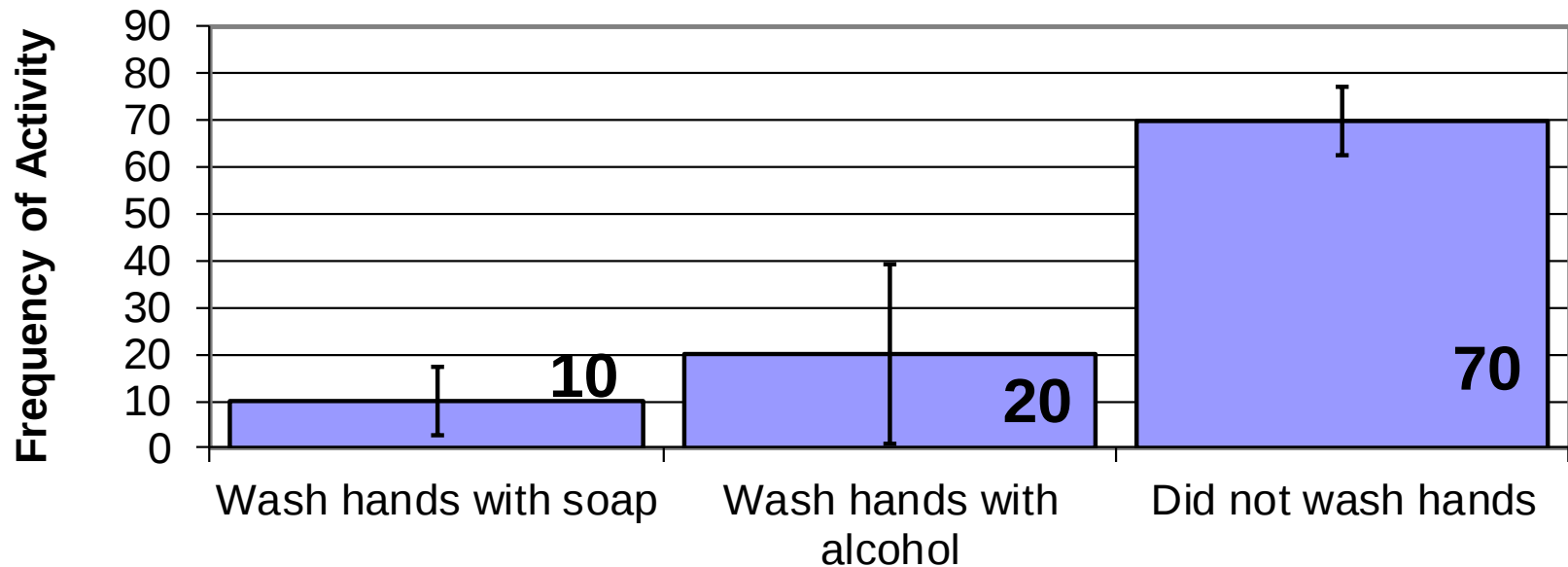
P-Value Interpretation

- Probability that the difference does not reflect a true difference and is only due to chance.
- e.g., $P=0.05$ means that 95 out of 100 times your estimate was truly significant
- Generally a level of $P<0.05$ is considered “statistically significant.”

95% Confidence Interval

Example:

Hand Hygiene Observations
in Hospital for 2010



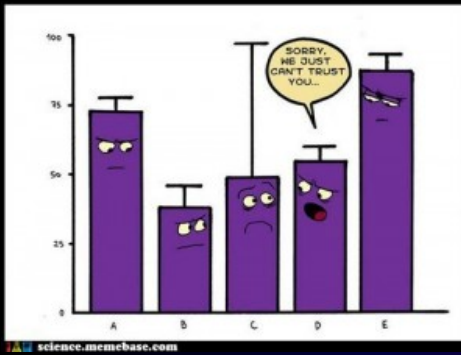
Statistical Variation of Estimates

- Consider your calculated infection rate to be an estimation of the true rate.

Why an estimation?

- You may only do surveillance on a sample of patients in your hospital.
- If surveillance activities were repeated by other ICPs, your numerators may vary slightly based on interpretation of case definitions, available clinical information in the chart, etc.

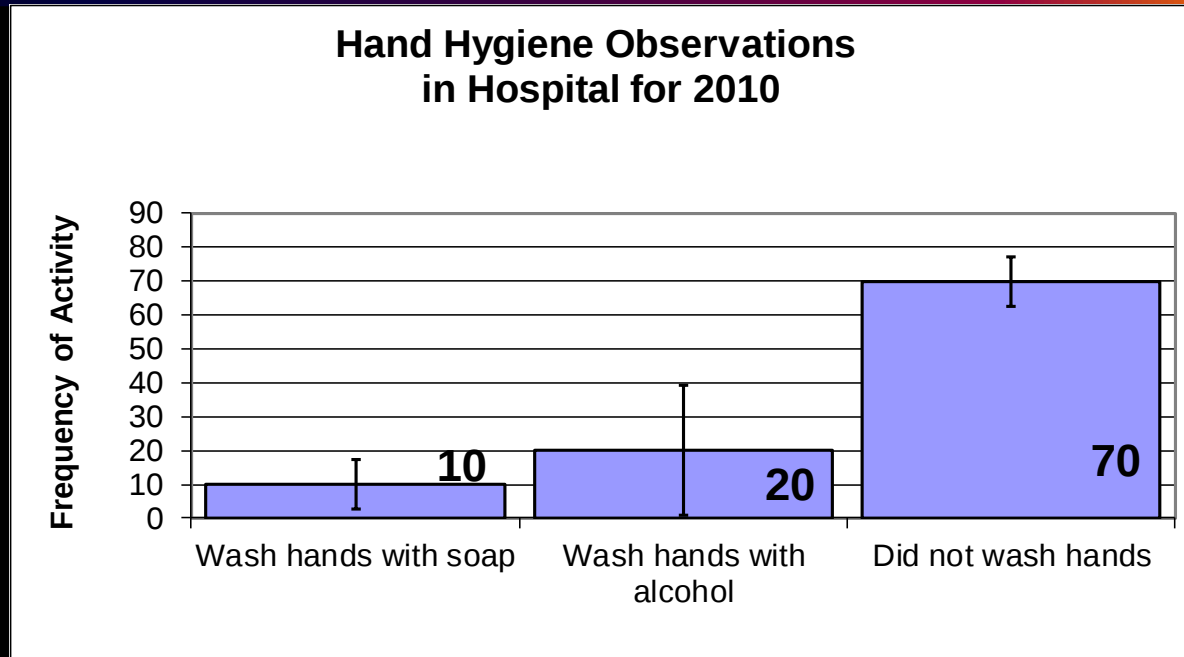
95% Confidence Interval Interpretation



- Means that you are 95% confident that the true average value lies within this interval.
- Confidence interval size:
 - Wide: less confident with that estimate
 - Narrow: more confident with that estimate
- For comparisons,
 - Overlapping intervals suggest no significant difference
 - Non-overlapping intervals suggest significant differences

95% Confidence Interval

Example:



Is the frequency of not washing hands at this hospital statistically significantly different than the frequency of washing hands with soap? **YES – the 95% CI do not overlap**

Is the frequency of washing hands with soap at this hospital statistically significantly different than the frequency of washing hands with alcohol? **NO – the 95% CI overlap**



Centers for Disease Control and Prevention
CDC 24/7: Saving Lives, Protecting People™

NHSN - National Healthcare Safety Network

NHSN Home

Alerts

Reporting Plan ▶

Patient ▶

Event ▶

Procedure ▶

Summary Data ▶

Import/Export

Surveys ▶

Analysis ▶

Users ▶

Facility ▶

Group ▶

Logout



Statistics Calculator

- ▶ [Compare Two Proportions](#)
- ▶ [Compare Single SIR to 1](#)
- ▶ [Compare Two Standardized Infection Ratios](#)
- ▶ [Compare Two Incidence Density Rates](#)
- ▶ [Compare Single Proportion to a Benchmark](#)
- ▶ [Compare Single SIR to Nominal Value](#)

Generate Data Sets

Reports

Statistics Calculator



Compare Two Proportions

When comparing two proportions (e.g. SSI Rates, Device Utilization ratios etc.), the hypothesis is that the rates are not different from each other. To perform a statistical test and calculate a p-value, enter the number of events as the numerator and the number of trials as the denominator (e.g. procedures, patient days) for two data sources. Press calculate.

	Data Source #1	Data Source #2
Group Labels:	2015	2016
Numerator (Number of Events):	2	10
Denominator (Number of Trials):	189	201

Title: ×

Calculate

Back

National Healthcare Safety Network

Colon Surgery SSI

As of: March 27, 2017 at 3:30 PM

	2015	2016
Numerator	2	10
Denominator	189	201
Proportion (shown as percentage)	1.1%	5.0%
Proportion p-value	0.0317	



Compare Two Incidence Density Rate

When comparing two incidence density rates (i.e. person-time), the hypothesis is that the rates are not different from each other. To perform a statistical test and calculate a p-value, enter the number of events as the numerator, the number of person-time units (i.e. exposure) as the denominator, and choose the multiplier you wish for the rate calculation. Press calculate. (See examples below)

	Data Source #1	Data Source #2
Group Labels:	Jan	Feb
Numerator(Number of events):	2	5
Denominator(Number of person-time units):	267	301
Multiplier:	1000 ▼	

Title:

Calculate

Back

Example 1

To compare 2 C.difficile LabID incidence rates:

- Enter the # of CDI HO Incident LabID events
- Enter the # of patient days
- Choose the desired multiplier(i.e., 10,000)
- Press calculate
- Output will provide the CDI HO Incident LabID Event rates per 10,000 patient days and the p-value to indicate the level of statistical significance

Example 2

To compare 2 Dialysis Event bloodstream infection rates:

- Enter the # of Dialysis Event positive blood cultures
- Enter the # of patient months
- Choose the desired multiplier(i.e., 100)
- Press calculate
- Output will provide the DE positive blood culture rates per 100 patient months and the p-value to indicate the level of statistical significance

Example 3

To compare 2 central-line associated bloodstream infection rates:

- Enter the number of CLABSIs
- Enter the # of central line days
- Choose the desired multiplier(i.e., 1000)
- Press calculate
- Output will provide the CLABSI rates per central line and the p-value to indicate level of statistical significance

National Healthcare Safety Network

Medicine ICU CLABSI Rate

As of: March 27, 2017 at 3:35 PM

	Jan	Feb
Numerator	2	5
Denominator	267	301
Incidence Density Rate	7.491	16.611
IDR p-value	0.3631	



Compare Single SIR to 1

When comparing a standardized infection ratio, the hypothesis is that the SIR is not different from one. To perform a hypothesis test and calculate a p-value, enter the number of observed events and the number of expected events. The SIR will be displayed automatically. Press calculate.

Data Source #1

Group Labels:

Number observed:

Number expected:

Standardized Infection Ratio: 0.714

Title:

Calculate

Back

National Healthcare Safety Network

ICU BSI Rate

As of: March 27, 2017 at 3:40 PM

Number Observed	Number Expected	SIR	SIR p-value	SIR95CI
5	7	0.714	0.4737	0.262, 1.583

Conclusions



- Describe Surveillance Data
- Display and Interpret Surveillance Data
- Determine the Significance of Changes to Surveillance Data

Questions?



Group Exercises Using Excel

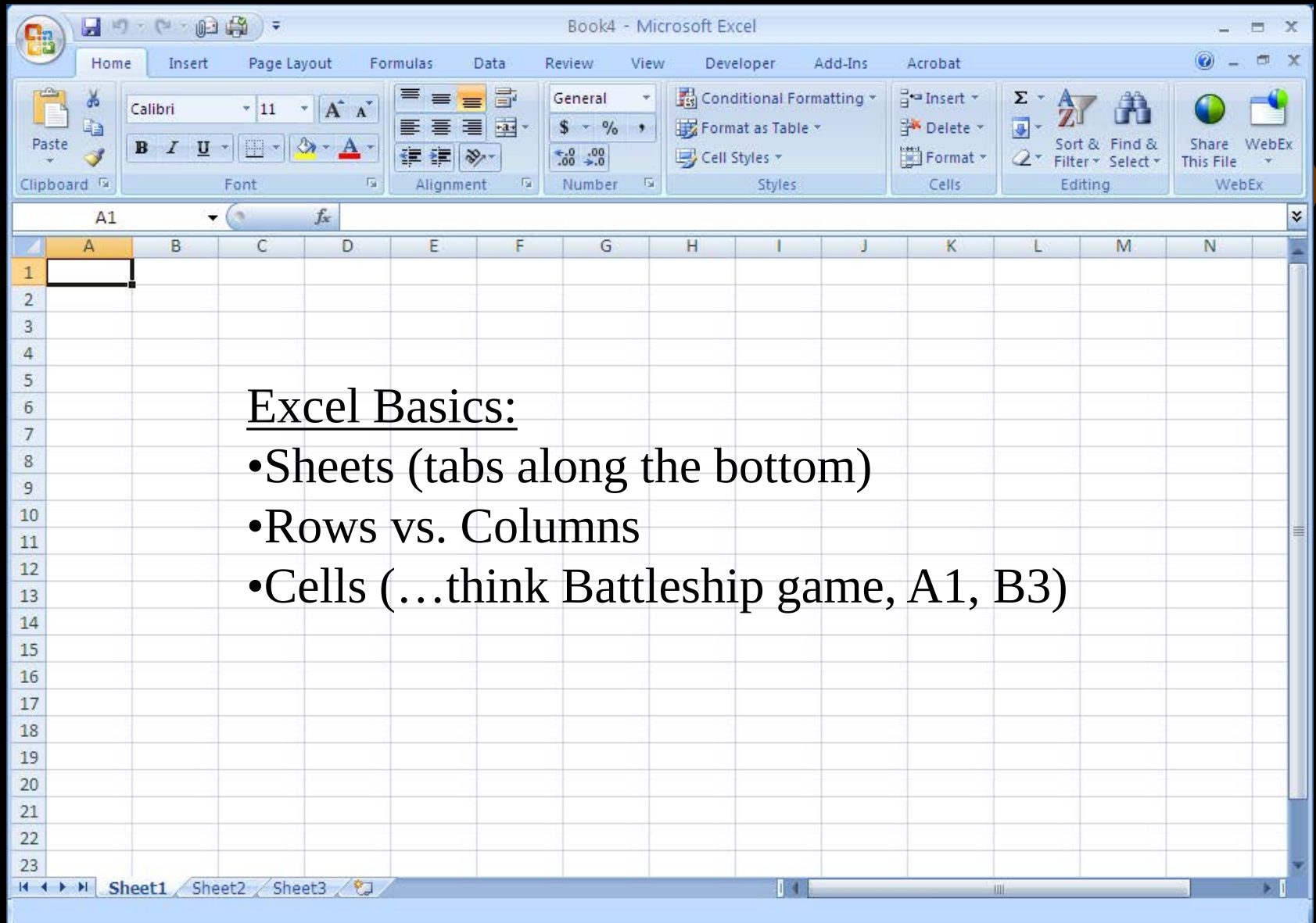


- Infection Rates
 - Create a table
 - Practice formulas
 - Optional activities
 - Graph rates
 - Add 2nd series on graph for NHSN benchmark
 - SIR calculation

Group Exercises Using Excel



- Outbreak Investigation
 - Create line-listing of outbreak cases
 - Practice formatting cells, copy/paste, sorting
 - Optional activities:
 - Create a frequency table of cases
 - Graph outbreak epi-curve



Excel Basics:

- Sheets (tabs along the bottom)
- Rows vs. Columns
- Cells (...think Battleship game, A1, B3)

Exercise Wrap-up



- Use Excel as a tool for
 - Calculations of infection rates
 - Creating line-listing for outbreaks or cluster investigations
 - Displaying data graphically
- Use each cell in Excel to capture single piece of data
- Graphs and tables should be self-explanatory!
 - Clear, concise title, informative labels
- Practice, practice, practice!