Slide numbers bolded and underlined: Answer slides completely bolded and italicized.

Slide 1: Introduction

Slide 2: These are also called **attribute** data (counted) – number of patients on a ventilator

and **variable** data (measured) – rate of vent associated pneumonia.

Slide 3: These will each be described in detail.

Slide 4: This is the crudest level of measure. There is no numerical value for each category. A value will clearly fit into only one category. An example of this could be a product trial of several brands of gloves. The choices would be brand A, brand B, brand C, or brand D.

Slide 5: These items are related to each other but the categories are distinct. They are ranked or scaled. Examples of this could be a survey with the choices of agree, neutral, disagree.

Slide 6: An example of this would be the MIC scale for antibiotic susceptibility.

Slide 7: Another example would be centigrade or Fahrenheit temperature scales

Slide 8: Answer: B- there is an assigned order to each value and the distance between each response is unknown.

Slide 9: Absolute measures are the simplest type of measurement, and they are also known as counts or frequencies. If you are measuring disease, then the number of cases would be an absolute measure of the amount of disease.

Slide 10: No notes

Slide 11: No notes

Slide 12: No Notes

Slide 13: Examples for each one

Slide 14: Number of people at risk—number of admissions or discharges to a unit during at time period.

Slide 15: Incidence should not be confused with **prevalence**, which is the proportion of cases in the population at a given time rather than rate of occurrence of new cases. Thus, **incidence** conveys information about the risk of contracting the disease, whereas **prevalence** indicates how widespread the disease is.

Slide 16: Answer is B. 5, The incidence rate for SSIs is 5/60 times a constant (probably 100 to yield a percentage) The two patients who were already there on June 1st are not part of the incidence.

Slide 17: No notes

Slide 18: Tell the participants after they have given you an answer, the first clue is that you can reject B. It was not written as a percent. This is simple division: 15/75 times 100. This is an example of the level of mathematical calculations that will be on the test. You can use a simple calculator, but you can also calculate these in your head. Answer: D.

Slide 19: This is another simple calculation. The ratio of total deaths to total population in a specified community or area over a specified period of time.

Crude mortality rate measures the proportion of the population dying each year from all causes.

Cause specific mortality rate measures mortality from a specified cause for a population.

The constant is usually 100,000 when you look at CDC statistics in the MMWR. (*In small countries, you may see 1000 as the constant because the total population is smaller. – not true*) This allows us to look at the crude mortality rate in one city, country or area and compare to another of a different size.

Slide 20: We don't have any information about deaths outside of ILI – we cannot answer this question Answer: E

Slide 21: The immediate clue for this answer is the "per 100,000". This is also a type of question that may be seen on the test. In this instance, the answer is 645 / 1,200,000 x 100,000 = 53.75/100,000 or about 54. Answer: A

Slide 22: Mean: the average value

– Useful for descriptive & inferential stats

– Measure of choice if data are normally distributed

• Median: the middle value; 50th percentile

– Useful for descriptive stats

– Measure of choice if data are skewed

• Mode: the most frequently occurring value

– Useful for descriptive stats

Slide 23: Standard deviation: square root of the sum of the differences between each value and the mean of the data set divided by the number of values in the data set minus 1

Slide 24: The range is 14 minus 2 = 12

Slide 25: No notes

Slide 26: No notes

Slide 27: This is an example of a question which could be on the test. The drawing of the bell curve won't be there. You will need to picture this in your mind using the data they give you. Using this data.

The SD is 1; The mean is 5.

Then you need to place the numbers on the chart knowing what the percents will be at -3, -2, -1, 0, +1, +2, and +3 SD from the mean.

Slide 28: Using the previous chart, what is the percentage of population between day 3 and 7?

Answer = B. 95.5% or they may give you the choices of: 68%, 96%, 99%, or 67%. You will need to pick the one nearest to the actual number.

Slide 29: The standard formula is the number of infections divided by the number of device days multiplied by 1000. The way to count the device days according to the NHSN system is to count the devices each day at a certain time (i.e., 2pm) Do not count all vents, central lines or catheters that were used during the 24-hour period. If a person is on the vent from 9am till 1pm and you count your days at 2pm, then that person would not be counted in the device day rate. But if that person develops an infection, he/she is counted (and meets all criteria).

Device utilization rates tell us how busy a facility is. When the DU rate is high, the activity at the facility is very high with higher risk patients. This tells the difference between a hospital which has very stable patients, such as a large obstetric hospital with low ventilator usage and a small or medium sized hospital with a lot of ventilators, central lines, etc. This number is important to show along beside the device associated infection rate.

Slide 30:   4/800   X 1000 is the same as   4000/800 which equals 5

The participants can use a simple calculator for the exam. Most calculations will be simple numbers which can be done in your head.

Slide 31: 800 / 4000 = 0.2

Slide 32: No Notes

Slide 33: No Notes

Slide 34: No Notes

Slide 35: No Notes

Slide 36: No Notes

Slide 37: SIR = 4 CLABSIs observed/0.50 CLABSIs predicted SIR=8

We observed more (8 times) CLABSIs than predicted based on comparison to a standard rate*

   *state source of standard rate, NHSN? which years?

Slide 38: Statistical significance is mathematical - it comes from the data (sample size) and from your confidence (how confident you want to be in your results). Practical significance is more subjective and is based on other factors like cost, requirements, program goals, etc.

Slide 39: Run charts are useful in identifying trends. They do not detect the difference between common cause and special cause variation. A run chart can be used with any type of data, discrete, continuous, nominal). Uses no statistical calculations – uses median of the data set.

Slide 40: SPC charts are more sensitive, use statistical calculations, data are plotted in time sequences, provide upper and lower control limits and helps to identify special cause.

When plotting the data on the chart, you need at least 16 data points for reliability. Follow the rules for evaluating a SPC.

There are several types of SPC charts – X and R (mean and range), X and s (mean and sample standard deviation), XmR (individuals' chart), c chart, u chart, p chart, np chart. See the APIC Text for a complete explanation of each type of control chart. Chapter 6 – Statistical Process Control.

Slide 41: **Common cause variation** is due to regular or normal causes and is relatively predictable.

Special cause variation occurs when an event or process is affected by an event outside the system. It is unstable because it is not predictable. It can be positive or negative. The source should be identified and eliminated.

An example of a positive or negative issue could be increased needlesticks due to change in products being used and decrease could be caused by low staffing and lack of time to fill out incident reports. Either one is an issue that needs to be identified and attacked. Based on

Gaussian theory of normal distribution, a special cause variation should be outside of 3 SD (99.73%) from the mean and should only occur 0.27% of the time.

Slide 42: Answer = A

Slide 43: When a study's results are shown to be statistically significant, it means that it is <u>highly unlikely</u> that the results were an accident. It doesn't mean that the results prove anything.

When not statistically significant, it could be an accident.

Slide 44: No Notes

Slide 45: An alpha error or Type I is rejecting a true null hypothesis. That means there was no difference between the antibiotic and the placebo, but you rejected that and said the research showed that there was a difference.

A beta error or Type II is when you say the null hypothesis is true - that there is no difference (between the effect of the pill and the placebo) but you are wrong. You should have rejected the null hypothesis.

Slide 46: The probability value (p-value) of a statistical hypothesis test is the probability of getting a value of the test statistic as extreme as or more extreme than that observed by chance alone, if the null hypothesis H0, is true.

The p-value helps determine rarity – how rare is this outcome that it could not have happened by chance alone.

Slide 47: The power of a study should be determined statistically so that you will know if you have a big enough sample of the population. You will not have to calculate the power of a research study population. The larger the sample size, the greater the accuracy of the results and we will have a better ability to apply the results to the general population. For example, if you want to know if a survey of 200

people is enough to tell you what the general consensus is in your community, you need to perform the statistical calculation.

Slide 48: Failing to reject a null hypothesis when you should is a Type II error or beta error. Answer = B. Type II error

Slide 49: Answer = A. 1 in 10

Slide 50: Answer = A

Slide 51: No notes

Slide 52: ANSWER B

Slide 53: Observational studies are those which review what has happened or what is happening without any manipulation by the observer. These can be descriptive – they merely describe what happened, when, where, etc.

Analytic studies usually show a comparison. The cross-sectional study shows everything at a specific point in time.

Experimental studies can be natural or planned but involve some manipulation by the person engineering the study. These will be described in detail.

Slide 54: In a Cohort study, the people start out well and healthy. Then they are followed over a specific time period. Then monitor all risk factors which have occurred in the people who developed disease and risk factors in those who did not develop the disease.

An example of this: You review all patients receiving a heart catheterization (they are healthy in that they usually don't have any skin infections). Then you check for risk factors such as poor hygiene, obesity, smoking, etc. At the end of 6 months, you have noted that 5 people have developed heart cath site infections. When you review the patients, you discover that most of those who have developed the infections are the very obese patients. So, you set up a 2 x 2 table to show the relationship of the obesity to the ones who became infected as compared to the obese patients who did not become infected.

Slide 55: The big disadvantage for the previous example is the large number of patients that had to be reviewed. You had to review all patients who received a heart cath. This takes an enormous amount of time. Your salary during the time spent reviewing all those charts is $$$$. And some patients may have developed infections and went to another doctor, so you may not know about all the results. These are said to be lost to follow-up. Some may have expired at home and no patient record of the event exists.

Slide 56: In these, you start with the results. (disease or a condition)

Using the same example, in a case control study, you start with the cath site infections. Then you select 5 matching or similar patients (or 10 matching) which did not develop the site infections. Then you review all patients to see which exposures occurred in the ones with infection and did not occur in those without infection.

You may have difficulty finding very similar patients to use as the control group.

Slide 57: No notes

Slide 58: This type of study is a snapshot picture at one point in time. The point prevalence is usually a specified time, such as, midnight on a specific day.

The period prevalence may be during the week of _____ or the month of ____. If a study does not indicate whether a point or period prevalence, you can usually assume it is a point prevalence study.

An example would be a point prevalence study of all patients with a central line in the facility at midnight. Then the patients would be analyzed as to a specific issue or risk factors.

Slide 59: No Notes

Slide 60: The important issue is that this study starts with healthy college freshmen and goes forward to observe them. It then divides the group into those who developed colds and those who did not. And then looks for risk factors (i.e., smoking). Answer: B. Cohort study

Slide 61: An example of this kind of study would be where one group is given a specific drug and the other group is given a placebo.

If the people who get the true drug are selected, then it is not random. If it is random, then a random selection process is used: every other person, every 4th person, etc.

In a double-blind study, the person who hands out the drug does not know which people get the real drug and which do not. There is documentation to know after the study is over who got the pill, but during the study the researcher does not know – so they can't cause the study to be affected and the results tainted.

Slide 62: Descriptive statistical methods include graphs, charts, and may be called Pictorial statistics.

Inferential statistics make an assumption about a population based on a small sample. It is used to show an association between cause and effect. Such as the association between the ventilator and pneumonia.

Slide 63: This is the basic 2 X 2 table. It can be set up with any 2 variables (other than disease or exposure).

A=exposed with disease

B=exposed but no disease

C=not exposed but has disease

D=not exposed and no disease

N is total population in study

Slide 64: This tells how strong the association is. This does not prove causation.

It is the incidence rate (number of new cases/population at risk) for those who are exposed divided by the incidence rate (new cases/population at risk) for those who have not been exposed.

Slide 65: If 4 out of the 14 who were exposed and only 1 out of those who were not exposed (11) became ill then your chances of getting the infection are higher if you are exposed. But how much?

0.29 divided by 0.09 equals 3.2.

So, you are 3.2 times more likely to get the disease if exposed, than if not exposed.

Slide 66: Another measure of association is the odds ratio.

The easiest example of this is to determine the probability of being a smoker if you have lung cancer divided by the probability of being a smoker if you do not have lung cancer.

If the odds ratio is equal to 1, the odds of disease are the same if the exposure is present and if absent (no association). If odds ratio is greater than 1, the odds of disease are higher for the exposed group and the exposure is probably associated with the disease.

Slide 67: No notes

Slide 68: Nine issues which strongly support the probability of causation

1. Strength - More people have lung cancer if they smoke.

2. Consistency - When you study smokers in different environments and areas, you come up with the same results.

3. Specificity – this relationship is only present with smoking and lung cancer and is not present with diabetes and lung cancer or obesity and lung cancer.

4. Time relationship – they only developed the lung cancer after they had been smokers for a while.

5. Biological gradient – the heavy smokers have higher rates of lung cancer than those who only smoke one pack a day.

Slide 69:

6. Plausibility – smoke can biologically cause lung cancer in humans

7. Coherence – those who smoke develop lung disease which progresses as they age and some of them then develop the lung cancer

8. Experiment – if it were ethically possible to do a double blinded or random study with one group smoking and the other not ever having smoked, and it resulted in the smokers developing lung cancer, it would carry a great deal of weight toward proof of causation.

9. Analogy – study shows that breathing harmful fumes other than cigarette smoke can cause lung damage and lead to lung cancer

All these are smoking guns which increase the chance that a causal association exists.

Slide 70: Non-parametric tests are used when your data is in the form of groups of simple data like: numbers of people with lice, scabies, or dermatitis are counted and tested. There will not be a normal "bell" curve for the data. It cannot be measured.

A parametric test would be used when you are measuring the blood pressure and temperature fluctuations of persons who are in surgery.

Your type of data determines the type of test you will use.

Slide 71: Z test – must have normal distribution.

Z test is frequently used to compare your SSI rates to those on the NHSN/NNIS charts. Their sample size is given on the chart for each type of surgery and is over 30. (Some statistics books [including the APIC Text] state that the denominator of either group you are comparing should be greater than 20 before you can use the Z test.) If the Z statistic is between -1.6 and 1.6 Standard Deviations and the data is normally distributed, the SSI rate is not significantly different from that of the standard population.

If the Z is less than -1.6 SD and case finding has been adequate, then the SSI rate is better than the standard population.

If the Z is greater than 1.6 SD then the SSI rate is significantly higher than the standard population and needs further work and investigation.

T test for smaller samples. Does not have to fit the bell curve.

Slide 72: In a one tail test, you will only be testing the hypothesis in one direction.

In a two tailed test, there are two regions where you can reject the hypothesis. With the new drug, you will want to know how many people are able to take the drug for the full regimen and you also want to know how many are only able to take the drug for a few days.

Slide 73: Non-parametric tests analyze 2 or more groups and measure the observed (the interest of the study) against the expected (the baseline or benchmark).

This is a test of statistical significance.

You will definitely need to know the difference between the Chi-square and the Fisher's exact tests for the certification exam.

Chi-square analysis (Pearson's) allows comparison of a set of data (observed values) to a set of theoretically generated values (expected values) in a formal way; observed and expected values are compared in terms of probability; the difference between these two values measures the agreement between the data and the theory, called a goodness of fit procedure Nonparametric.

Fisher's exact- tests an association between two variables; nonparametric; use when expected 2x2 table data are less than 5 (instead of the Chi-square); The two categories must be mutually exclusive and independent.

Slide 74: In an outbreak, the Chi-square test is used to evaluate the probability that observed differences between 2 populations (such as cases and controls) could have occurred by chance alone if an exposure is not truly associated with disease. There are several variations of the chi-square test.

A copy of the chi-square table is in any statistics book.

Slide 75: Both of these tests should be done by computer software program to obtain the accurate P-value. Despite the fact that the formula for this test calculates a P-value, you must calculate the P-value for the observations and then add this to the values of all possible combinations that have lower P-values. Use a computer!

Slide 76: Answer= D. you need to know more information to choose the test.

Slide 77: No notes

Slide 78: Example: the standard HIV screening test (ELISA) has a 99.5% sensitivity when it comes to testing positive for the presence of HIV antibodies. This means that if you have 1000 people who have HIV and you test them with this test, 995 will test positive and 5 will have a false negative.

Slide 79: The specificity is the accuracy of the negative result (in those who do not have disease). If you have a 95% specificity, then you will have 5% who will falsely test positive.

When you change a test to make the sensitivity higher, the specificity will be lower. (and the reverse is true).

Slide 80: Probability that patients with a positive test truly have the disease

Probability that subjects with a negative screening test truly don't have the disease.

PPV ask "If the test result is positive, what is the probability that the patient actually has the disease?"

NPV asks "If the test result is negative, what is the probability that the patient does not have disease?"

Slide 81: Answer: Sensitivity is 40/50 X 100 = 80%

Specificity is 70/100 X 100 = 70%

Slide 82: Bar graphs are used to compare facts. The bars provide a visual display for comparing quantities in different categories or groups. Bar graphs help us to see relationships quickly. However, bar graphs can be difficult to read accurately. A change in the scale in a bar graph may alter one's visual perception of the data.

Line graphs are used to display data or information that changes continuously over time. Line graphs allow us to see overall trends such as an increase or decrease in data over time.

Circle Graphs are used to compare the parts of a whole. Circle graphs represent data visually in the same proportion as the numerical data in a table: The area of each sector in a circle graph is in the same proportion to the whole circle as each item is to the total value in the table. Constructing an accurate circle graph is difficult to do without a computer. This is because you must first find each part of the whole through several elaborate calculations and then use a protractor to draw each angle. This leaves a lot of room for human error. Circle graphs are best used for displaying data when there are no more than five or six sectors, and when the values of each sector are different. Otherwise, they can be difficult to read and understand.

Slide 83: No Notes

Slide 84: Bar chart compares the size and magnitude of differences – bars are not connected since data is mutually exclusive

Slide 85: No Notes

Slide 86: No Notes

Slide 87: Histogram is used to graph a frequency distribution of a set of continuous data. A continuous data set consists of a series of measurements for which there is an infinite number of possible values between the lowest value and the highest value in the set. Columns should be adjoining with the height of each column being proportional to the frequency of events in that interval. Shows the number of cases over time.

For more impact, place labels and arrows at points where interventions were done, or actions happened with a note or words to describe. This chart could represent the needlesticks during implementation of a safety program.

MOST often used in an outbreak investigation

Slide 88: Use: a bar graph that focuses on critical issues by ranking them from the most frequent to the least frequent.

Helps to identify the order in which to start problem resolution.

Helps to monitor the success of a project.

Data depicts that 80% of the problems come from 20% of the individuals or events.

Slide 89: Answer: C. Pareto chart.

Slide 90: No Notes

Slide 91: No Notes

Slide 92: Answer: B. Point-source epidemic likely people exposed at same source over a brief time, such as through single meal or single event