Epidemiology Principles
Surveillance
Descriptive stats

Utah CIC Study Group
9/20/2022

Week 11   9/20/2022

Epidemiology, surveillance, performance and patient safety measures

Group discussion

  General principles of epidemiology (Study guide ch 4)

  Surveillance (Study guide ch 4)

  Descriptive statistics (Study guide ch 4)

Independent review:

  Outbreak investigations (Study guide ch 4)

Week 12   9/27/2022

Epidemiology, surveillance, performance and patient safety measures (cont)

Group discussion

  Process control charts (Study guide ch 4)

  Risk-adjusted comparisons (Study guide ch 4)

  Quality concepts (Study guide ch 4, 7)

Independent review: (None)

Week 13   10/4/2022

Epidemiology, surveillance, performance and patient safety measures (cont)

Group discussion

  Performance measures (Study guide ch 4, 7)

  Qualitative research methods (Study guide ch 4)

  Research study design (Study guide ch 4)

Independent review:

  Patient safety (Study guide ch 4, 7)

Complete chapter 4 study guide practice questions

Questions – Diseases
Certification Study Guide, 6th ED

Thanks: From the Kerns River APIC chapter

- **Questions – Epidemiology**

- Ch 3: 11

- Ch 4: 1, 2, 5, 6, 10 – 38, 41, 42, 44, 45, 47, 50

- Ch. 5: 44

- Ch. 8: 8, 11, 13, 14, 16,

- Exam 1: 2, 10, 24, 26, 29, 32,34, 35,37, 45, 53, 60, 61, 68, 72, 77, 80, 84, 90, 99, 107, 108, 112, 113, 116, 125, 132, 135

- Exam 2: 4, 6, 12, 17, 39, 40, 45, 47, 48, 57, 60, 69, 70, 74, 80, 84, 85, 88, 102, 104, 110, 116, 118, 119, 124, 128, 129, 131, 132

- Exam 3: 4, 8, 10, 11, 16, 18, 26, 41, 46, 48, 49, 56, 57, 67, 70, 77, 91, 94, 102, 104, 107, 113, 124, 127, 131

- For more certification resources, please visit www.apic.org/store and www.apic.org/education

- The primary purpose of epidemiology is to aid in the understanding of the causes of a disease by knowing its distribution; determinants in terms of person, place, and time; and natural history.
- Understanding the elements involved in the transmission of infection enables infection preventionists (IPs) to develop strategies that target specific areas in the transmission process.
- Selecting the appropriate study design is an essential step in answering questions important to IPs.
- IPs should understand the meaning of commonly used epidemiology terms and know how to apply basic epidemiology skills.
- Correct presentation of data allows the IP to demonstrate outcomes and relationships in a manner that will likely encourage collaboration and support among stakeholders.
- Risk: The probability or likelihood of an event occurring.
- Risk factor: A characteristic, behavior, or experience that increases the probability of developing a negative health status (e.g., disease, infection).
- association: The relationship between a risk factor and an outcome, such as a disease.
- Causation: A determination that a particular risk factor results in an outcome, such as a disease.
- Epidemiology: The study of the distribution and determinants of disease and other conditions in human populations.
- Hypothesis: A theory developed with current information regarding a risk factor or set of risk factors, which can be tested using statistics.

Write or Draw ⬆ Memory

Sink
Fish
M&M+m
Chai
Web
Table
SpIN SnOUT

Surveillance Ch 11

Calculated incidence or prevalence (in the now, or previous to now)

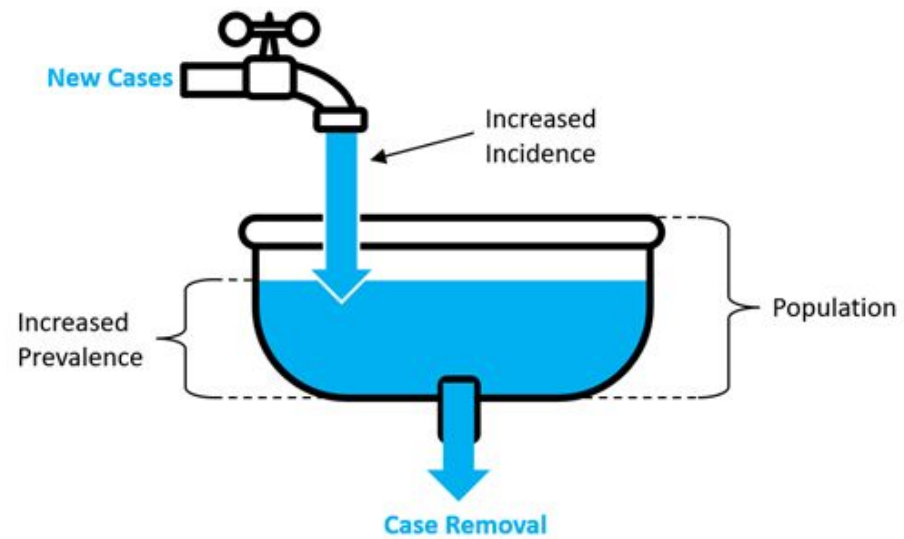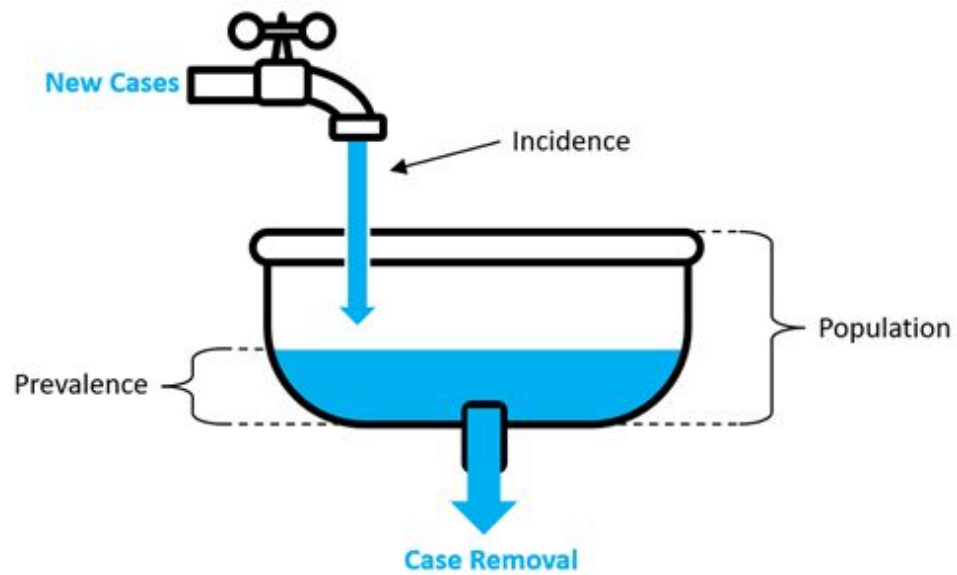Calculated specific infection rates /ratios

Provider specific, unit specific, device specific, procedure specific, standardized infection ratio

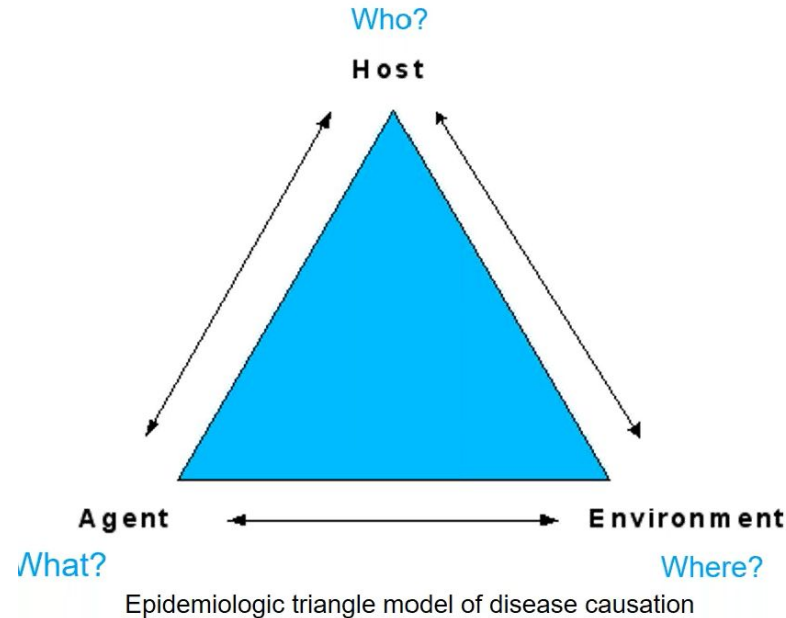Interpretation of surveillance data (descriptive statistics)

Syndromic surveillance (symptoms before the lab result usually)

Sentinel Surveillance (often larger places with an enhanced surveillance like flu)

Pseudo outbreaks (cohorts coming in for school from another country with bulk testing?)
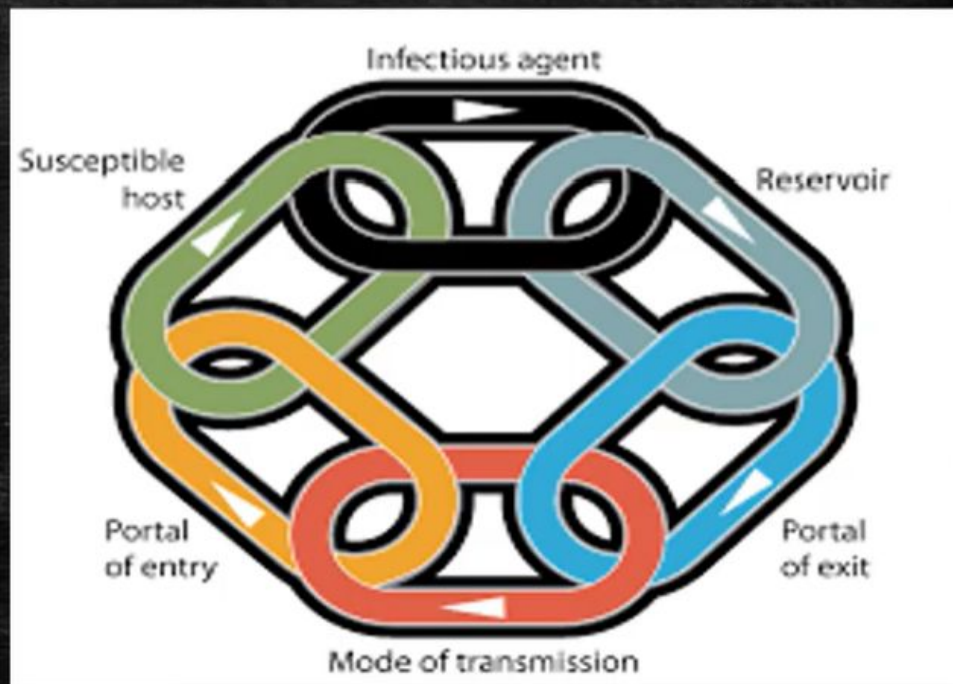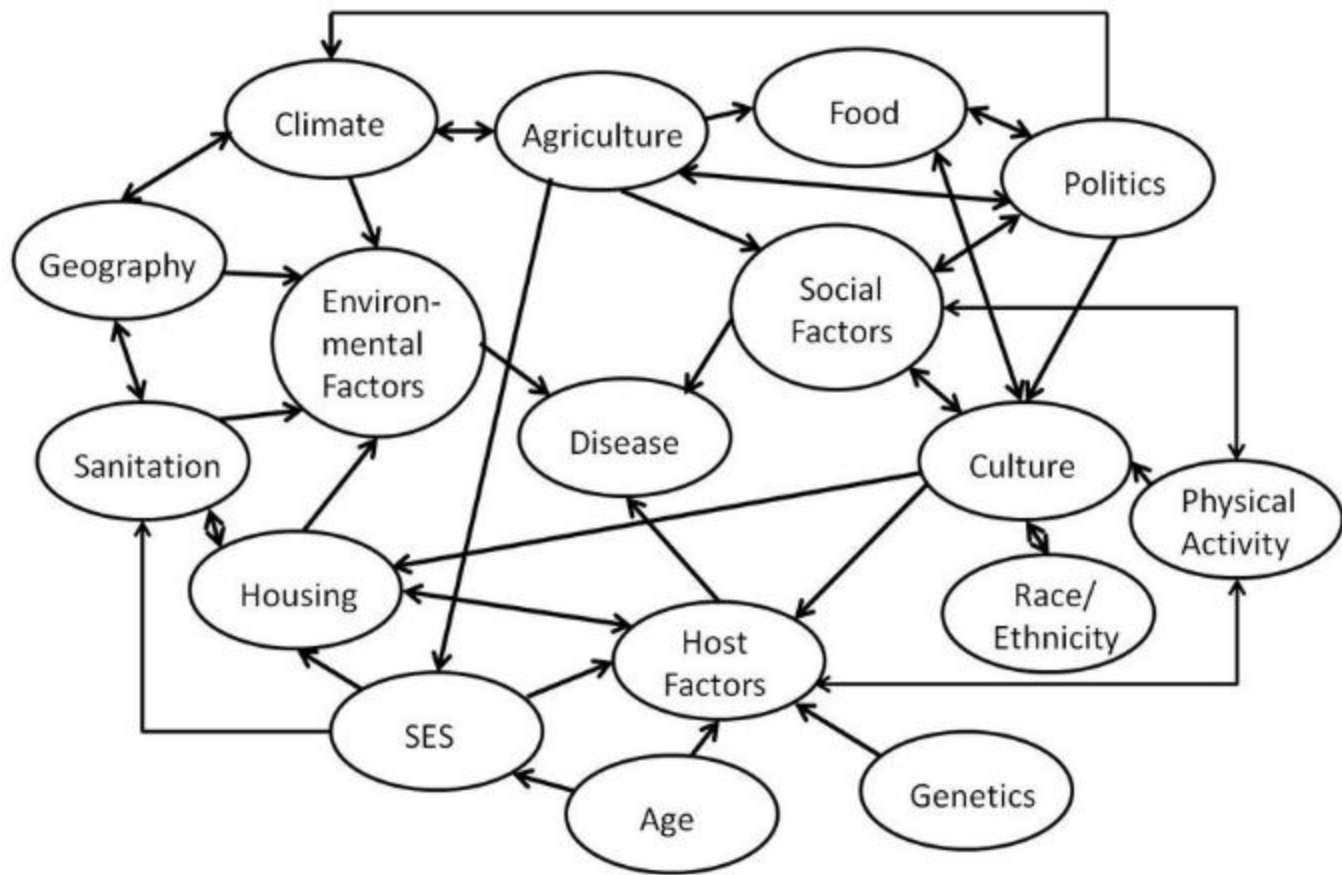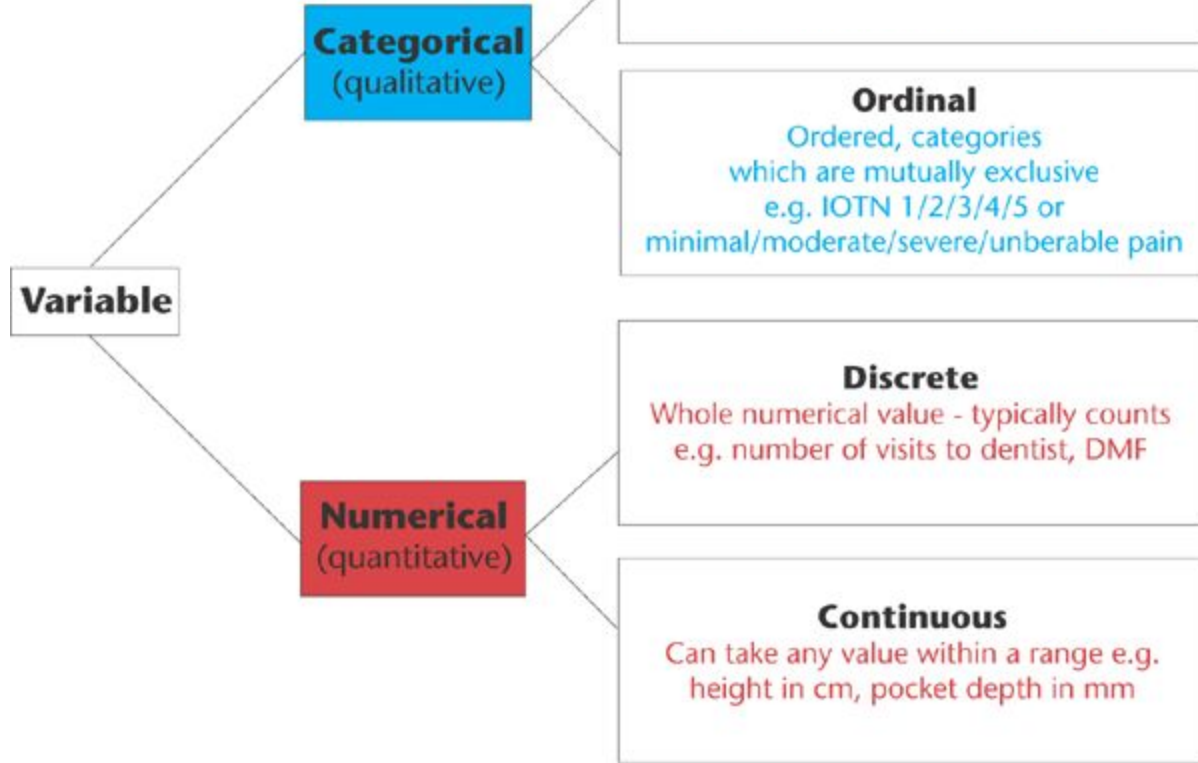
Chapter 13 of APIC text

Who?

Host

Agent ↔ **Environment**

What? Where?

Epidemiologic triangle model of disease causation

Epidemiology – Basics

- Chain of Infection — only have to break 1 link to stop an infection

So many variables!

Variable

Categorical (qualitative)

Nominal
Unordered, categories
which are mutually exclusive

Ordinal
Ordered, categories
which are mutually exclusive
e.g. IOTN 1/2/3/4/5 or
minimal/moderate/severe/unberable pain

Numerical (quantitative)

Discrete
Whole numerical value - typically counts
e.g. number of visits to dentist, DMF

Continuous
Can take any value within a range e.g.
height in cm, pocket depth in mm

Relationship between Measures of Central Tendency under different types of Skewness

NORMAL DISTRIBUTION & THE EMPIRICAL RULE

A major reason to examine histograms is to decide whether measures of central tendency represent the dataset well. If the histogram has a single, central peak and is roughly symmetrical (i.e., most data values are near the center and the more extreme values balance each other), a measure of central tendency is appropriate. The ideal version of this profile, in which the mean, median and mode are identical, is called the normal distribution (Figure 13-4).

Figure 13-4.
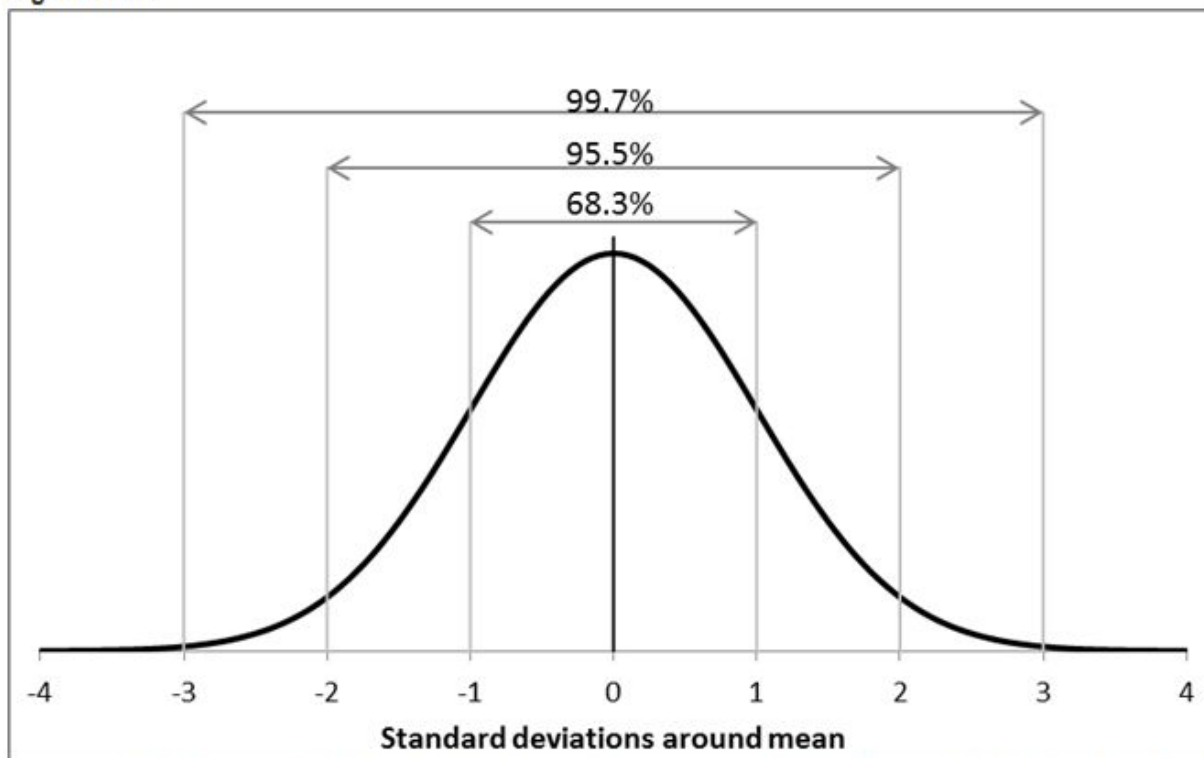
A normal distribution around a mean of 0, and percentages of data found within one, two or three standard deviations of the mean

Levels of Prevention

Onset

Clinical diagnosis

No Disease

Asymptomatic Disease
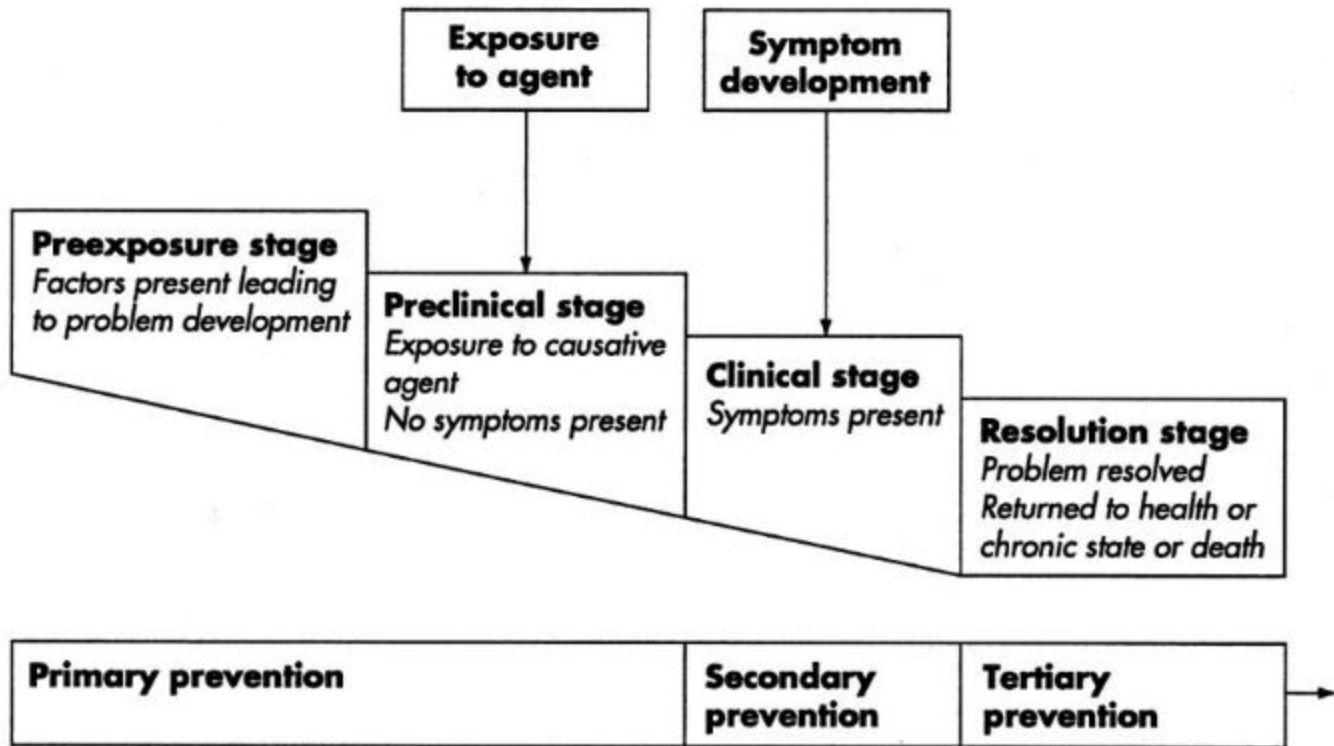
Clinical course

Primary

Secondary

Tertiary

Goal: the complete prevention of a disease before any manifestation of that disease occurs

(e.g., wellness programs, immunizations, folic acid, exercise, seat belts)

Early diagnosis and treatment and preventing further deterioration by intervention as early in the disease course as possible

(e.g., screening, skin testing in TB, mammograms,)

Reducing complications

(e.g., rehabilitation and organ transplantation)

Exposure to agent

Symptom development

Preexposure stage
Factors present leading to problem development

Preclinical stage
Exposure to causative agent
No symptoms present

Clinical stage
Symptoms present

Resolution stage
Problem resolved
Returned to health or chronic state or death

Primary prevention

Secondary prevention

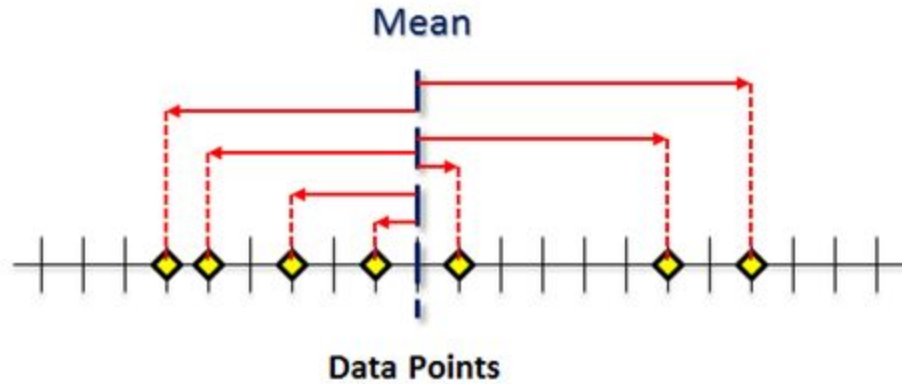**Tertiary prevention**

M&M' +M

Mean-average

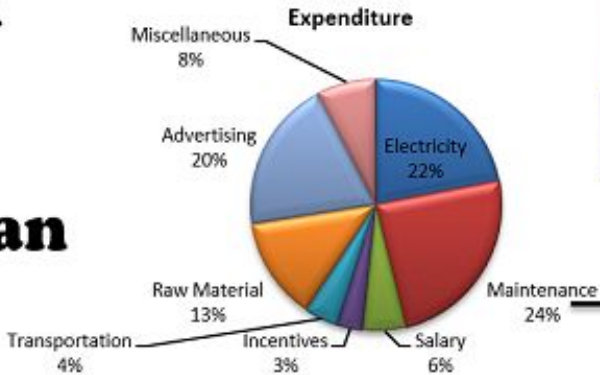Median-half above, half below

Mode-Most frequent

Standard deviation is a summary measure of how far the data points spread out around the mean

A (very) little statistics overview
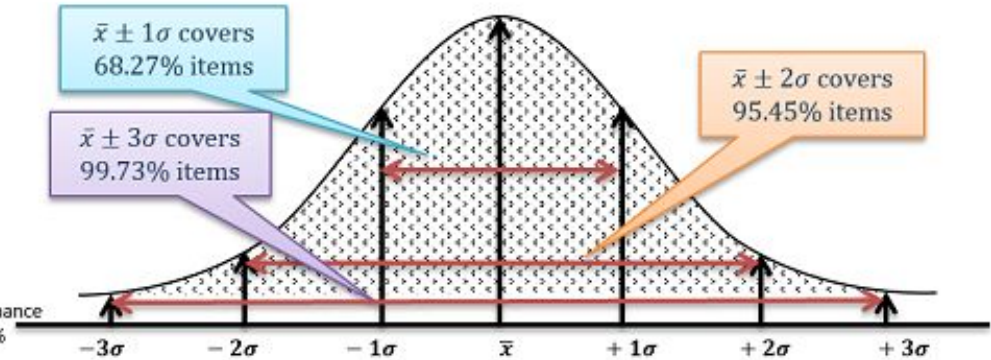
Descriptive Statistics

Mean

Median

Mode

$$Std.\,Dev.\,\sigma = \sqrt{\dfrac{\Sigma(x - \overline{x})^2}{n}}$$

Expenditure

Miscellaneous 8%
Advertising 20%
Electricity 22%
Raw Material 13%
Transportation 4%
Incentives 3%
Salary 6%
Maintenance 24%

$\overline{x} \pm 1\sigma$ covers 68.27% items

$\overline{x} \pm 3\sigma$ covers 99.73% items

$\overline{x} \pm 2\sigma$ covers 95.45% items

-3σ -2σ -1σ $\overline{x}$ $+1\sigma$ $+2\sigma$ $+3\sigma$

Measures of Central Tendency: Mean, median, and mode
Measures of Dispersion (spread): Standard deviation, variance, range, quartiles
Normal distribution: symmetric around the mean

Correlation (r): direction and magnitude of a relationship between 2 variables

- Positive correlation (+)r = as one variable increases, so does the other

- Negative correlation (-) r= as one variable increases, the other decreases

- No correlation (o)r= no association between the two variables

p value= Power. The probability of obtaining a statistic at least as large as what you would have gotten if the hull hypothesis (Ho) were true.

Distribution:

- Range– highest value minus the lowest

- Deviation– actual distance of each observation from the mean

- Standard deviation– measures the average spread around the mean

- Variance– standard deviation squared

Parametric tests for association: Normally distributed data

- More accurate– likely to reject Ho when it is false. Only valid with normal distribution.

- Z test– compares means between different samples

- Student's t test—compares means between 2 different samples

    - One-tailed or two-tailed

- ANOVA (comparing means between > 2 samples)

- Regression analysis

Non Parametric tests for association: Non-normally distributed data

- Chi square

    - Discrete data (observed against expected)

    - Measures observed against expected

    - Cannot use if one cell is 2 x 2 table is < 5

- Fischer's exact text used for small sample size

- Mann -Whitney U test (or Wilcoxon rank sum and Kendall's S) compares median between 2 samples

Chi-square

discrete data

Fisher's test for small #

N=< 20

< 5 per cell

Small fish

Sensitivity= likelihood that someone with the disease/condition will test positive (greater sensitivity means fewer false negatives). a/a+c

Specificity= the likelihood that someone that does not have the disease or condition will have a negative test result (higher spec ificity means fewer false positives). d/b+d

Positive Predictive Value= the likelihood that someone who tests positive actually has the disease or condition. a/a+b

Negative Predictive Value= the likelihood that someone who tests negative actually does not have the disease or condition.

True (+)	False (+)
c	d
False (-)	True (-)

Hypothesis = Ho

Hypothesis/ Conclusion	True	False
Accept Ho	No error	Type II error (beta)
Reject Ho	Type I error (alpha)	No error POWER

Figure 13-21.
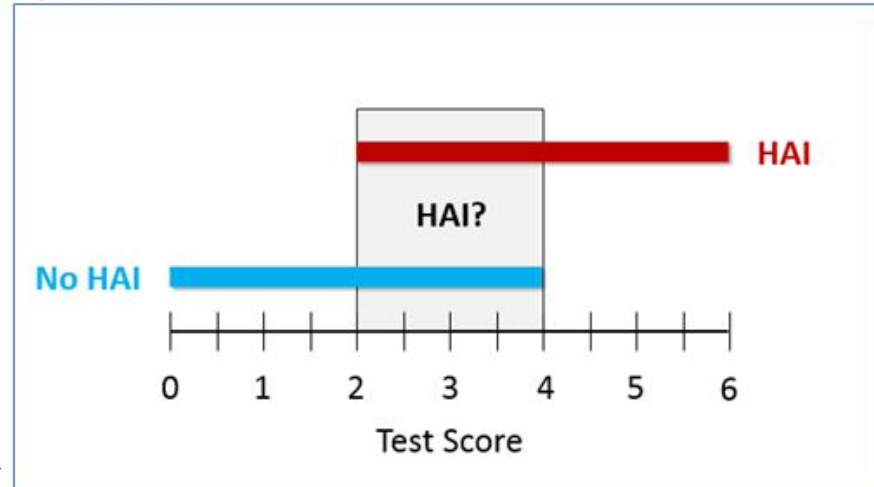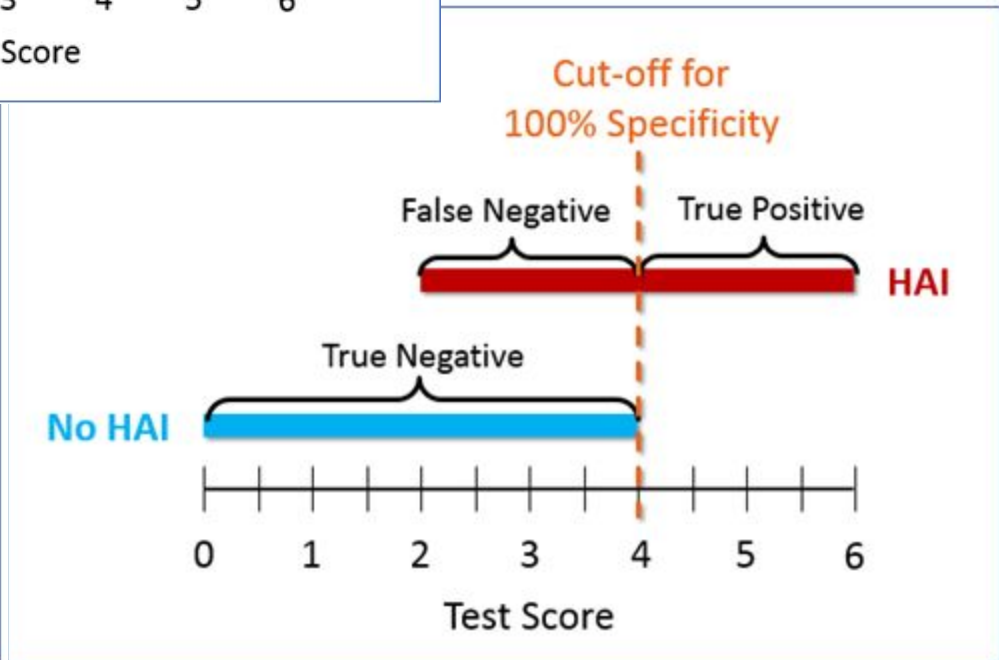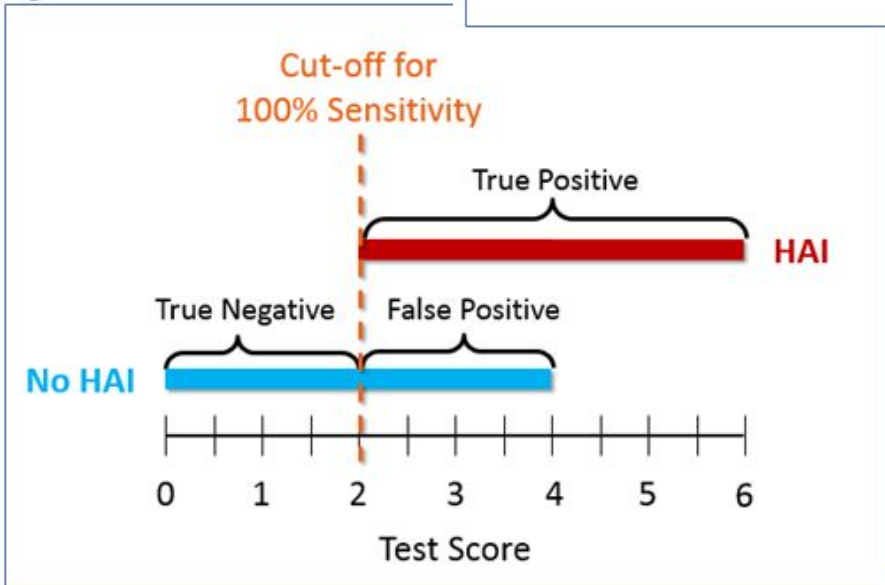
Figure 13-22.

Many of the following definitions are adapted or borrowed from the glossary in *Principles of Epidemiology in Public Health Practice*.[1]

- **2 by 2 table:** a table with two outcome columns (e.g., disease and no disease) and two exposure rows (e.g., exposed and not exposed). See Table 13-1.
- **Categorical data:** variables with data split into mutually exclusive categories (also known as qualitative data).
- **Denominator:** the lower portion of a fraction; used in calculating ratios, proportions, and rates.
- **Quantitative data:** variables with numeric data that represents counts or values on a numeric scale.
- **Incidence:** a measure of the frequency with which new cases or events occur among a population during a specified period.
- **Numerator:** the upper portion of a fraction.
- **Population:** the total number of people in a particular group.
- **Prevalence:** the number or proportion of cases or events present in a given population.
- **Ratio:** a comparison of two quantities, calculated by dividing one quantity by the other.
- **Sensitivity:** the ability of a test or other surveillance method to identify true cases.
- **Specificity:** the ability of a test or other surveillance method to exclude persons who are not cases.
- **Validity:** the degree to which a screening test or other data collection tool measures what it is intended to measure.
- **Variable:** any characteristic or attribute that can be measured and can have different values.

Incidence Prevalence Endemic Epidemic Pandemic Outbreak

The number of existent cases of a given disease at a given time	**Prevalence**
An excess over the expected incidence of disease within a given geographical area during a specified time period	**Epidemic**
An epidemic spread over a wide geographical area, across countries or continents	**Pandemic**
The usual incidence of a given disease within a geographical area during a specified time period	**Endemic**
Is a sudden increase in occurrences of a disease in a particular time and place	**Outbreak**
The number of new cases of a given disease in a given time period	**Incidence**

Florida HEALTH

Reservoir | Fomite | Herd immunity | Risk | Risk Factor

Definition	Term
The probability or likelihood of an event occurring	**Risk**
A place in which an infectious agent can survive but may or may not multiply	**Reservoir**
A characteristic, behavior, or experience that increases the probability of developing a negative health status	**Risk Factor**
An inanimate object on which organisms may exist for some period of time	**Fomite**
The resistance of a group to invasion and to spread of an infectious agent, based on the immunity of a high proportion of individual members of the group	**Herd immunity**

Florida HEALTH

Epidemiology = epi (upon) + demos (people) + logos (study).

Descriptive epidemiology = amount and distribution of disease within a population. Precedes analytic epidemiology. Uses numbers to tell a story.

Analytic epidemiology = reasons for frequency of disease.

Inferential epidemiology = makes an assumption about a population from a sample. Hypothesis testing.

Prevalence = existing cases at a given time.

Incidence = development of new disease over a certain time period.

Incidence rate = new disease cases divided by the population at risk.

Incidence density = new disease cases among persons at risk during a specific time.

Attack rates = ill ÷ ill + well x 100 during a time period; Noted as a %. Exposed and ill.

Infectiousness = # infected divided by the # susceptible.

Pathogenicity = # with disease divided by the # infected.

2x2 table compares those with disease to those exposed. Can show odds ratio or relative risk.

Odds Ratio = a x d/ b x c= . Is the odds of exposure in cases divided by the odds of exposure in controls (no disease).

Relative Risk = a/a+b divided by c/c+d. Is the rate of disease in exposed divided by the rate of disease in unexposed.

Rate Ratio = Incidence density in exposed divided by the incidence density in the unexposed. Can use a 2 x 2 table and substitute person or time for the No Disease column.

	Disease	No Disease	Total
Exposed	a	b	a + b
Not Exposed	c	d	c + d
Total	a + c	b + d	N

Bradford-Hill's Criteria for Association

Criteria	Explanation
1. Strength of association	The stronger the relation between a risk factor and the effect (outcome), the less likely it is that the relation is due to a third or extraneous factor.
2. Consistency	Multiple studies in a range of settings report similar results.
3. Specificity	Ideally, the effect has only one cause.
4. Temporality	The purported cause should be present before the effect occurs.
5. Biological gradient	A dose–response relation between the risk factor and the effect.
6. Biological plausibility	There should be a rational and theoretical basis explaining how or why the risk factor led to the effect.
7. Coherence	The association should not conflict with known facts.
8. Experimental evidence	Is there any supportive research based on experiment; if preventive action is taken, does the effect dissipate?
9. Analogy	A previously accepted phenomenon in one area can be applied to another.

Additional resources:
Determining Causality: A Review of The Bradford Hill Criteria
https://www.youtube.com/watch?v=TnuosYuKGos
Bradford Hill causal criteria
https://www.youtube.com/watch?v=-VFk-Pg6Yjo
Bradford Hill Criteria (UNC)
https://www.youtube.com/watch?v=2h8u_WCAcYc

Florida
HEALTH

Three Types of Association

Artificial (Spurious)

May be caused by :

Errors in study design or analysis

Bias

Errors in analysis

Failure to control for confounding variables

Indirect or Noncausal

May be caused by:

Mixing of effects between the exposure, disease and confounding variable that may be associated with the exposure and independently affect the outcome of interest

Causal

Occurs when: evidence indicates that one factor is clearly shown to increase the probability of the occurrence of a disease

Do NOT confuse with causality!

Florida HEALTH

Healthcare-associated Infection

Community-acquired Infection

Not present at the time of admission to the hospital but are temporally associated with admission to or a procedure performed in a healthcare facility

An infection POA may also be healthcare-associated if it is related to a recent hospitalization
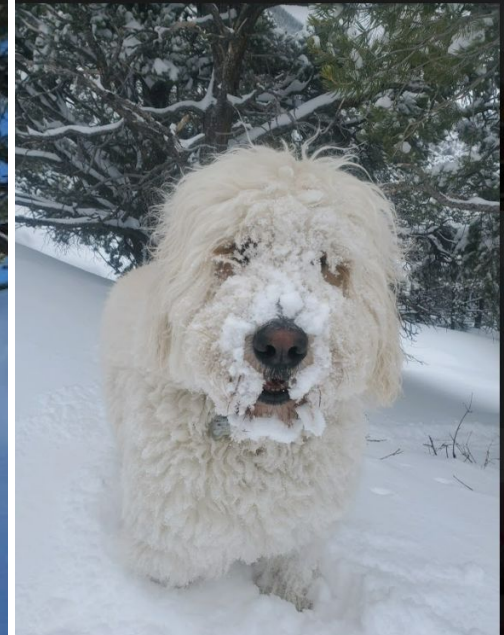
Infections present on admission with no association to a recent hospitalization

Watch for bias like rose colored glasses and head in the sand

Questions

Identify the median in the following list of numbers: 6, 2, 9, 7, 1, 4

    a.   9

    b.   7

    c.   5

    d.   4

Question #22

In any normal distribution, the proportion of observations that are within two standard deviations of the mean is closest to:

a. 0.50
b. 0.68
c. 0.95
d. 0.98

Question #23

The most important feature of nonparametric tests is that they:

a. Make no assumption about variance in the populations
b. Can only be used with ordinal levels of measurements
c. Require normal distribution
d. Require equal population variances

Question #23, Answer/Rationale

Answer: A, Make no assumption about variance in the populations

Rationale: Non-parametric data make no assumption about the distribution of the population values and can be used with discrete, nominal, ordinal, or interval data.

The main advantage of non-parametric methods are that assumptions of a normal distribution are NOT required, however, they may not provide accurate results because they are distribution-free.

Question #24

Which of the following indicates a strong positive correlation?

a. $r = 0$
b. $r = -0.993$
c. $r = 0.603$
d. $r = 0.45$

Question #24, Answer/Rationale

Answer: C, r=0.603

Rationale: Correlation shows whether paired variables are related. The r value measures the degree of strength of the relationship. The range of r values is -1to 1. The closer the value is to 1, the stronger the positive relationship, meaning as one variable increases, so does the other. A negative correlation means that as one variable increases, the other decreases.

Note that correlation is not the same as causation!

Question #26

Which of the following steps are NOT included in hypothesis testing?

a. State the null and alternative hypothesis
b. Set the significance level
c. Eliminate outliers
d. Compare the probability value to the significance level

Question #26 Answer/Rationale

Answer: C, Eliminate outliers

Rationale: Hypothesis testing is a key part of statistical analysis. It estimates the probability (likelihood) that a result did not occur by chance.

Steps to hypothesis testing include:

1. State the question
2. Specify null and alternative hypotheses
3. Calculate test statistic (comparison to evaluate relationship to null, e.g. t-test)
4. Compute probability of test statistic (p-value)
5. State conclusions

Outliers are observations that deviate significantly from other observed values. These must be investigated to determine their origin and the optimal method of analysis, however, simple removal of outliers can drastically change conclusions.
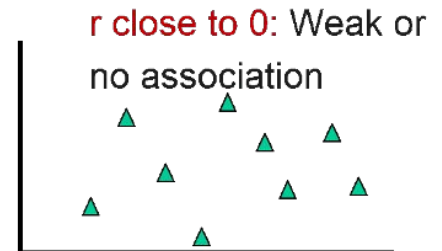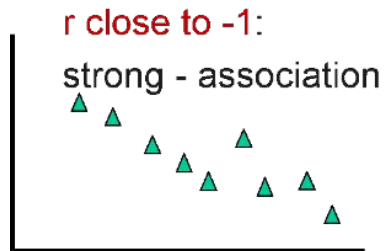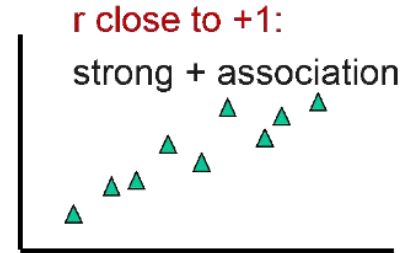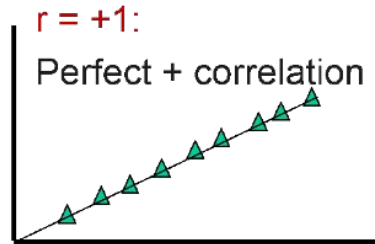
Question #27

The range of the correlation coefficient is:

a. -1 to 0
b. 0 to 1
c. -1 to 1
d. None of the above

Question #27, Answer/Rationale

Answer: C, -1 to 1

Rationale: Correlation is used to evaluate the direction and magnitude of the relationship between two variables. The closer the value is to +,- 1, the stronger the correlation. The association decreases as r approaches zero.
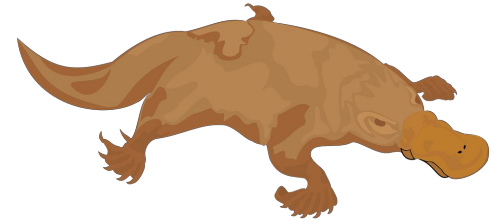
Question #28

If the index of kurtosis is -1.99, then the curve is:

a. Relatively flat
b. Negatively skewed
c. More peaked
d. A typical bell shaped curve

Question #28, Answer/Rationale

Answer: A, Relatively flat

Rationale: Kurtosis is a statistical measure used to describe the degree to which scores cluster in the tails or the peak of a frequency distribution. Positive kurtosis indicates a high peak with "skinny" tails and negative kurtosis indicates a flat distribution with "fat" tails. A normal distribution has an index of zero.
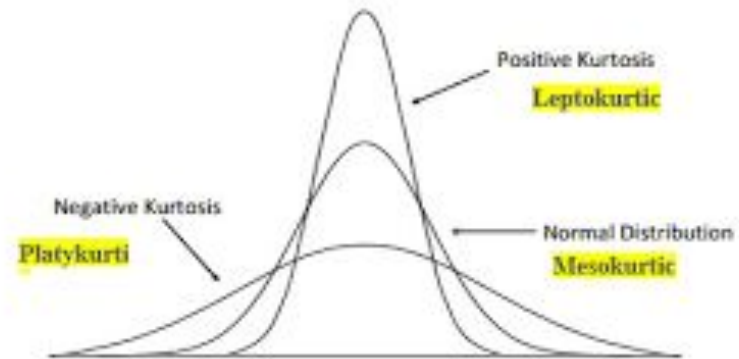
Question 1

A causal association is:

1. A false association that can be due to chance or bias in study methods
2. Mixing of effects among exposures, disease and a third factor that is associated with the exposure and independently affects the outcome
3. When evidence indicates that one factor is clearly shown to increase the probability of occurrence of a disease
4. Determined through a set of criteria including strength of identification, consistency and biological plausibility

   A. 1 and 2
   B. 1 and 3
   C. 3 and 4
   D. 2 and 4

Question 2

Which of the following is an example of the criterion of "Strength of Association" from Hill's criteria for causation?

A. In a study of the association between abx exposure and development of *C.difficile* infection, the odds ratio was 2:3

B. In a study of the association between abx exposure and development of *C.difficile* infection, the author's conclusion are consistent with those of three other studies

C. In a study of the association between abx exposure and development of *C.difficile* infection, antibiotic therapy began an average of 3 weeks before *C.difficile* infection

D. In a study of the association between abx exposure and development of *C.difficile* infection, prolonged antibiotic therapy was a greater risk factor for *C.difficile* infection

Question 3

- The medical director in one of your facility's outpatient dialysis units is requesting your assistance in investigating three cases of Clostridium difficile (C. difficile) infection. All three patients received their treatments on the same day, but at different times. You agree to assist with the investigation. An outbreak is defined as:

1.   infection or event that occurs at a rate higher than expected

2.   When an unusual microbe is recognized

3.   An incidence rate that is more than two standard deviations above usual

4.   The identification of three or more cases
   - A.   1,4
   - B.   3,4
   - C.   1,2
   - D.   2,3

Question 4

The "epidemiologic triangle" model for disease causation does *not* include:

A. Agent

B. Host

C. Time

D. Environment

Question 5

Which of the following are considered primary prevention measures:

1. Barrier precautions
2. Immunization of HCP
3. Cleaning and disinfection
4. Sterilization
    A. 1,2
    B. 3,4
    C. 1,2,3
    D. 1,2,3,4

CIC Statistics Knowledge and Skills

- Explain the different types of quantitative data
- Explain the different types of categorical data
- Describe the various measures of central tendency
- Describe the various measures of spread
- Explain the difference between descriptive and inferential statistics
- Describe the various measures of frequency

- Define null hypothesis
- Define alternative hypothesis
- Explain the concept of p-value
- Explain reliability, sensitivity, and specificity
- Differentiate between positive and negative predictive value
- Explain what "percentiles" mean
- Define what the Standardized Infection Ratio (SIR) reports
- Generate a SIR

Question 3

An acute care facility experiences an outbreak of *Serratia marcescens* bloodstream infections. After the outbreak is under control and no new cases are being reported, the IP wants to find the source of the outbreak. The most appropriate epidemiology study design to use is:

A. Retrospective cohort study

B. Prospective cohort study

C. Case-control study

D. Cross-sectional study

Rationale Q3

C- case control studies group by disease status and then investigate past exposures with objective of identifying exposures that are most common to cases than to controls. This is an appropriate study design for this example because there are existing cases, and the IP is trying to identify the exposures that are associated with the bloodstream infections.

Question #25

The Employee Health Service has notified the IP that seven employees have *P. aeruginosa* folliculitis. Initial investigation reveals that six of the seven cases belong to the same health club. Working on the hypothesis that the whirlpool at the health club is associated with the infections, the IP decides to conduct a case-control study using two controls for each case. Which of the following groups is the most appropriate control.

a. Non-ill family members of the ill employees
b. Non-ill hospital employees matched for age and sex
c. Hospitalized patients with *P. aeruginosa* folliculitis matched for age and sex
d. Non-ill members of the health club matched for age and sex.

Angel	Eyeball	Pizza
Angry	Fireworks	Pumpkin
Baby	Flower	Rainbow
Beard	Flying saucer	Recycle
Bible	Giraffe	Sand castle
Bikini	Glasses	Snowflake
Book	High heel	Stairs
Bucket	Ice cream cone	Starfish
Bumble bee	Igloo	Strawberry
Butterfly	Lady bug	Sun
Camera	Lamp	Tire

NEXT WEEK HAS BEEN EXAUSTING