

Foundations in Biostatistics and Epidemiology

Module 1: Epidemiologic Measures of Disease Burden and Distribution

The following provides a brief summary of the material that is included in the session modules.

Part I: Descriptive Statistics

Terminology:

- Population describes the hypothetical (and usually) large number of people to whom you wish to generalize
- Sample describes those individuals who are in the study (fraction of the population); utilize a sample of data to draw inference to the population of interest

Types of data:

- Qualitative: provides qualitative attributes
 - Dichotomous or binary: observations fall into one of two categories
 - Nominal: no obvious ordering of the categories
 - Ordinal: natural ordering of the categories
- Quantitative: quantifies amounts, numeric data
 - Discrete: observations can take on certain numeric values (e.g., counts) characterized by gaps
 - Continuous: not restricted to take on only certain values, often measurements

Common descriptive statistics:

- Qualitative data:
 - Counts, proportions, percentages are used to summarize distributions
- Quantitative data:
 - Measures of location or central tendency
 - Mean
 - Median (50th percentile)
 - Relation: mean=median for symmetric distributions; mean > median for positively skewed distributions; mean < median for negatively skewed distributions
 - Measure of spread or variability
 - Variance, standard deviation (s.d.): How much variability can we expect among individual responses?
 - Standard error of the mean (s.e.m.): How much variability can we expect in the mean response among various sample of size n?
 - Interquartile range: 75th percentile minus 25th percentile
 - Range: maximum minus minimum

$$s.e.m. = \frac{s.d.}{\sqrt{n}}$$

Part II: Descriptive Plots (see slides for a description of the plots)

Part III: Epidemiologic Measures of Disease Burden and Spread

Terminology:

- Count - # of cases of health event
- Ratio: Division of one quantity by another ; most general term; units and range varies
- Proportion: A ratio in which the numerator is contained in the denominator; indicates fraction of the population that is affected; no units; range 0 - 1
- Rate: a ratio representing change over time; range 0 to infinity
- Odds: Defined as the probability of an event (P) divided by the probability of a non-event (1-P); range 0 to infinity
- Prevalent cases: Existing cases (old and new) at a specified point in time or over a period of time

- Prevalence: not a measure of risk; used to express “burden of disease” used by health planners for determining workload; for monitoring control programs for chronic conditions
- Incident cases: New cases identified over a specified period of time
 - Incidence: tells us about change in status from non-diseased to diseased (limited to new cases); etiology => direct indicator of risk of disease (cumulative incidence); comparing incidence rates in populations
- Prevalence Proportion: the proportion of the population that has the disease at a given time point or period

$$\text{Prevalence Proportion} = \frac{\text{\# of existing cases of a disease at a given point in time}}{\text{Total population}} \times 1,000$$

- Point prevalence: numerator and denominator reflect a point in time

$$\text{Point prevalence} = \frac{\text{\# of existing cases of a disease at a point in time}}{\text{Total population at same point in time}} \times 1,000$$

- Period prevalence: numerator and denominator reflect a time period

$$\text{Period prevalence} = \frac{\text{\# of existing cases of a disease during a period of time}}{\text{Average population during the same period of time}} \times 1,000$$

- Cumulative Incidence (Risk): the proportion of initially susceptible individuals in a population who develop new cases of a disease in a specified time period

$$\text{CI} = \frac{\text{\# of new cases of a disease during a given period of time}}{\text{Population at risk during same time period}} \times 1,000$$

- Incidence Rate (also called Incidence Density) – the number of new cases divided by the person-time of observation; useful when individuals are observed for different lengths of time

$$\text{ID} = \frac{\text{\# of new cases during the time period}}{\text{Total person-time at risk during study period}} \times 1,000$$

- Person-time - an estimate of the actual time at risk in years, months, or days that all persons contributed to the study
- Relation between prevalence and incidence: in steady state and with low prevalence disease: expressed mathematically as Prevalence= Incidence x Duration

Part IV: Epidemiologic Measures of Mortality

Terminology:

- Mortality rate = incidence of death in a population; Number of deaths occurring in a specified population in a given time period
- Crude mortality: total death rate in an entire population (generally per 100,000 person-years)
 - # of deaths / total population for a given year
- Cause-specific mortality: rate at which deaths occur for a specific cause
 - # of deaths from specific cause / total pop for a given year
- Age-specific mortality
 - # deaths for age group/ total pop in age group for a given year
- Proportionate mortality ratio: Useful for identifying leading causes of death; Gives the relative importance of a specific cause of death in relation to all deaths

$$\text{PMR} = \frac{\text{\# of deaths from given cause in specified time period}}{\text{Total deaths in same time period}} \text{ Per 100}$$

- Case fatality ratio: Refers to proportion of fatal cases among those who have the disease; Provides an index of the deadliness of a particular disease within a specific population

$$\text{CFR} = \frac{\text{\# of deaths due to disease X}}{\text{Number of cases of disease X}} \times 100$$