July 21 - 24, 2021 | Garki Hospital Abuja | Resource Person

Research Methodology Boot Camp

with Epi Info Training

Dr. Adamu Onu
MBBS, FWACP (FM)
MS Epidemiology & Biostatistics
PhD Public Health (Epidemiology)

Target Audience Clinical Researchers, Post-Part 1 Residents, and Others

Important Information

- Limited slots are available on a first come, first served basis
- Laptop running Windows 10 required
- Organized as morning lecture sessions and afternoon hands on coaching sessions

Highlights

- Research Methodology
- Research Design
- Data Management
- Sample Size Calculations
- Test Statistics
- Interpretation of Results
- Report Writing
- Hands-on training sessions
- Statistical consulting sessions

For further details contact
Email: epimetrix@gmail.com
Phone: +234 803 474 9930

epimetrix
health systems
data for health decisions

Descriptive statistics

Introduction

- Data accumulated during research will need to be summarized and presented in a clear accurate format.
- For this tools of ***descriptive statistics*** are used.
- After summarizing and describing the results, the next step is to ***analyze*** the results.
- The results of the study must be clearly summarized to allow for proper analysis and interpretation

Parameters vs. Statistics

Parameters

- Numerical descriptive measures corresponding to *populations*.
- Unknown constants referred to in Greek letters (μ, π, σ)
- Statistical inferential methods based on **sample data** can be used to make statements (or inferences) concerning the **unknown** parameters.

Statistics

- Numerical descriptive measures corresponding to *samples*.
- Samples are ‘random subsets’ of the population, so statistics are random variables — different samples will yield different values of the statistic.
- Referred to in lowercase letters ($\bar{x}$, $\hat{p}$, s).

Descriptive statisitcs

- Descriptive statistics helps us make sense of a large volume of data.
- The following tools can be used in describing and summarizing the results:
  - Tabulation,
  - Calculation,
  - Graphs/figures
  - Correlation

Tabulation

- A first step in summarizing the data is commonly to group them in summary tables.
- The plan for the table should be developed in the research design phase.
- The term dummy tables is used to describe tables that are not yet filled with data.
    - Frequency tables
    - Cross-tabulation tables

Frequency distribution table

- A frequency distribution table gives the frequency with which a particular value appears in the data.
- Suitable class boundaries are needed.
- Classes must also be mutually exclusive.
- For example, if we tabulate data about the diastolic blood pressure, we may make the classes as 70 to 79, and 80 to 89, not 70 to 80 and 80 to 90
- Frequency distribution tables describe one variable at a time.

Cross tabulations

- Depending on the study objectives, the relationship between several variables may be examined at once in order to look for differences or relevant associations.
- Cross tabulation tables may be **descriptive** or **analytical**.
  - Descriptive cross-tabulations may be used to describe the sample
  - Analytical cross-tabulations may be used to determine differences between groups.
- The accepted convention is to put the categories of the dependent variable as column headers, and the independent variable as row headings.

Calculations

- Numerical data can be summarized by calculating their:
  - Central tendency
  - Variability
  - Percentage and proportions
  - Ratios and rates.

Central tendency

Mean

- Derived by summing up the individual values and dividing by the total number of measurements:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Central tendency

Median

- The median of a distribution is a midpoint at which one half of the observations fall below abd one half fall above the value

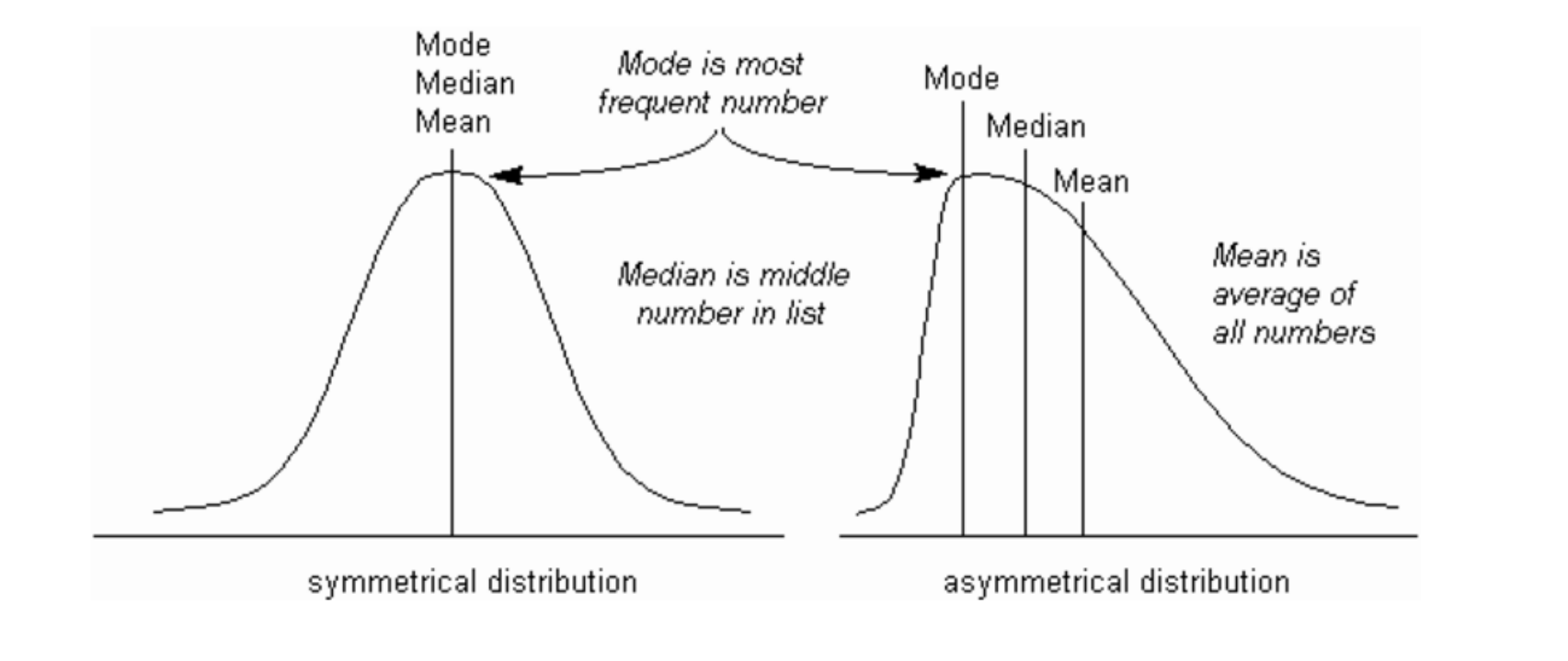
$$m(x) = \begin{cases} x_{\frac{n+1}{2}} & \text{if } n \text{ is even} \\ \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n+1}{2}}\right) & \text{if } n \text{ is odd} \end{cases}$$

Central tendency

Mode

- The mode is the most frequent measurement.

- In a “normal” distribution, the mean, median and mode coincide.
- In “skewed” distributions they vary and may all be meaningful in the presentation of the data.



Variability

- It is important to have some idea about the variation of the data around the mean.
- **Range**
  - The range gives the minimum & maximum values, but does not give much indication of the spread of observations around the mean.
- **Standard deviation** (SD)
  - The standard deviation is a measure of the spread of observations around the mean.

Percentages

- Usually missing data are not included in the calculation of percentages.
- Caution should be exercised when describing percentages based on small numbers.
- In such cases, a small difference may appear as a big difference in percentages.

Proportions, ratios & rates

- Proportion
  - A numerical expression that compares one part of the study units to the whole. Can be expressed as a fraction (2/5) or a decimal (0.40)
- Ratio
  - A numerical expression of the relationship between one set of frequencies and another. An example is the ratio of males to females in a sample.
- Rate
  - A numerical expression of the frequency of a condition in a given population measured in a specified period of time.

Graphs/Figures

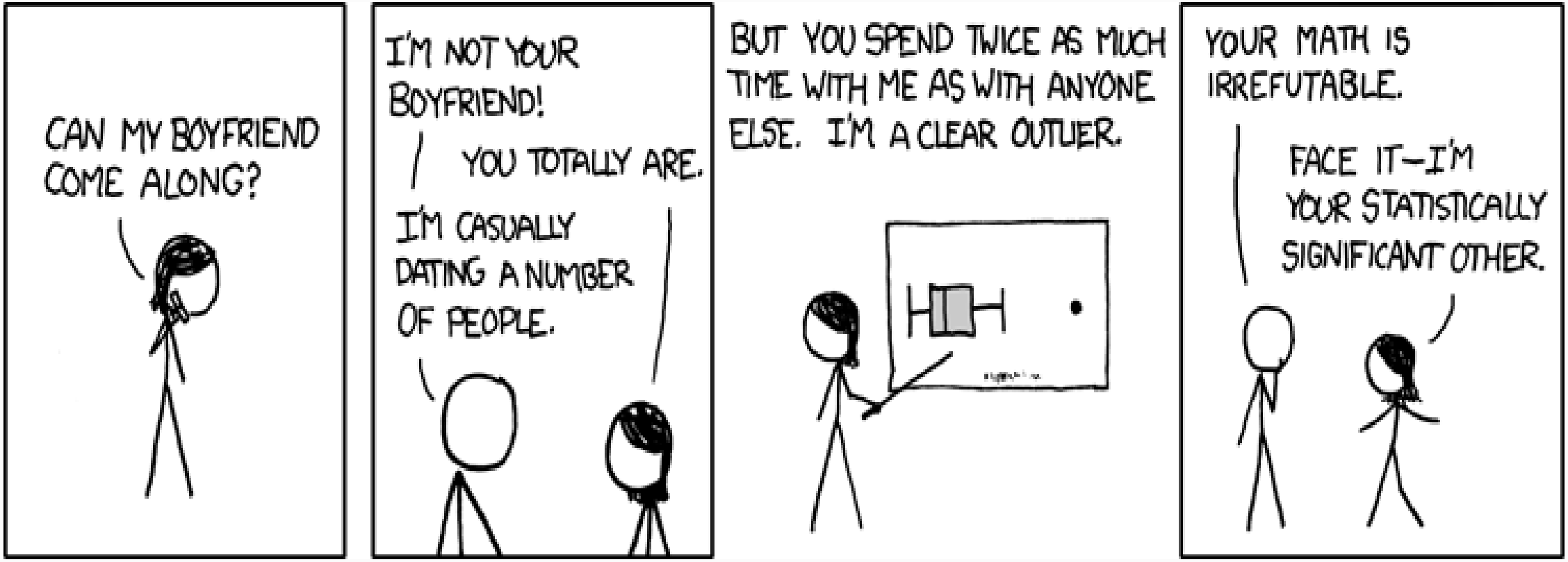
- Bar charts
- Pie charts
- Histograms
- A histogram resembles a bar graph but the bars are drawn to touch each other, reflecting the underlying continuity of the data.
- Line graphs
- Maps

Scatterplots

- When we want to see if there is a relation between two sets of observations
- The scatter diagram gives an indication whether a correlation may exist and its direction.
- The independent variables are graphed on the x-axis and dependent variables are graphed on the y-axis.

Examples – salaries dataset

Gender	Salary	Age	Place	Weight
Female	1500	33	Chicago	80
Female	1200	33	Chicago	82.5
Male	2200	34	New York	100.8
Male	2100	42	New York	90
Female	1500	29	Chicago	67
Female	1700	19	Washington	60
Male	3000	50	Washington	77
Male	3000	55	Washington	77
Female	2800	31	New York	87
Male	2900	46	New York	70
Female	2780	36	Washington	57
Male	2550	48	New York	64


CAN MY BOYFRIEND COME ALONG?
I'M NOT YOUR BOYFRIEND!
YOU TOTALLY ARE.
I'M CASUALLY DATING A NUMBER OF PEOPLE.
BUT YOU SPEND TWICE AS MUCH TIME WITH ME AS WITH ANYONE ELSE. I'M A CLEAR OUTLIER.
YOUR MATH IS IRREFUTABLE.
FACE IT—I'M YOUR STATISTICALLY SIGNIFICANT OTHER.