UNIVERSITY OF HAWAI'I
Office of Biostatistics & Quantitative Health Sciences
JOHN A. BURNS SCHOOL OF MEDICINE

Research Design & Biostatistics

Lecture 1

John J. Chen, Ph.D.
Professor
Office of Biostatistics & Quantitative Health Sciences
UH JABSOM

Surgery Resident Academic Session
October 17, 2018

Lecture Note: http://biostat.jabsom.hawaii.edu/Education/training.html

Outline

Lecture 1 (10/17/2018)

- The goal of statistics
- Introduction to descriptive biostatistics
- Some research design and data presentation issues

Lecture 2 (December, 2018)

- Large databases

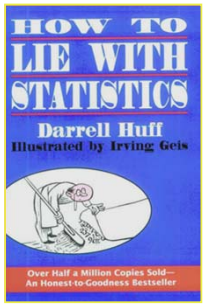
Lecture 3 (January, 2019)

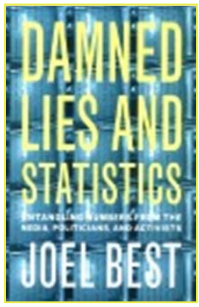
- Introduction to inferential statistics
- Some commonly used statistical approaches

Lecture Objectives

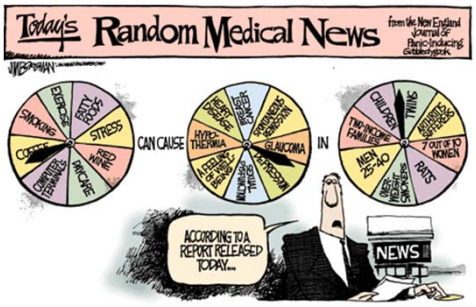
- Clarify the goal of statistics
- Grasp descriptive statistics
- Be familiar with various data presentation approaches
- Introduce key concepts of inferential statistics
- Survey some commonly used statistical approaches
- Understand basic research design principles
- Understand the pros and cons of large databases
- Build a foundation which will facilitate the active participation in clinical and translational research

Lies, Damned Lies, And Statistics





Medical Research, Media, and Public Health



Definition of Statistics

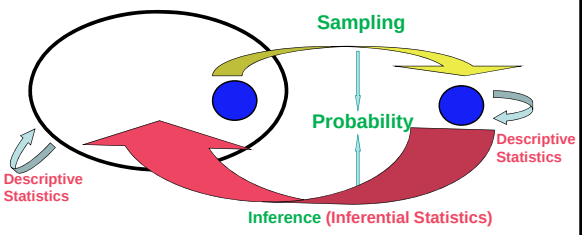
The theory and methodology for research (study) design, and for describing, analyzing, and interpreting information (data) generated from such studies, in which the data is subject to chance variation.

Population & Sample

- *Population*: the set of all subjects of interest having a common observable characteristic. For example, all patients who had heart surgery in US.
- *Sample*: a subset of a population, e.g., all heart surgery patients at QMC in 2017.

- *Parameter*: a summary measure of the population, e.g., the average LOS of the above population.
- *Statistic*: a summary measure of the sample, e.g., the average LOS of the above sample.

The Goal of Statistics



Population
[Parameters]
(μ, σ)

Sample
[Statistics]
($\bar{X}$, S)

Properties of A "Good" Sample

- Adequate sample size (statistical power)
- Random selection (representative)

Commonly used sampling techniques

1. Simple random sample
2. Stratified sample
3. Systematic sample
4. Cluster sample
5. Convenience sample

Types of Data & Scales of Measurement

1. Qualitative variables - categorical

- Nominal: Categories, names (e.g., gender, eye color)
- Ordinal: Ordered data, intervals are not equal (e.g., satisfaction scores, grades of tumor)

2. Quantitative variables - numerical

- Discrete - no intermediate values (e.g., number of children per family)
- Continuous – intermediate values (e.g., temperature, birth weight)

Types of Variables

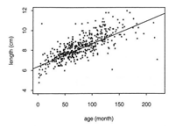
Notes:

Dependent (response) versus

Independent (explanatory) variables

In linear regression analysis:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$



Sources of Data (Types of Studies)

Two major types of investigations:
Surveys versus experiments

Major difference: whether the investigator has control over which subjects enter each study group.

Some examples of survey researches
Prospective (cohort) studies
Retrospective (case-control) studies
Cross-sectional studies

Some examples of experimental studies:
Lab experiments
Clinical trials

Descriptive Statistics

Qualitative data:

- Frequencies
- Percentages

Quantitative data:

- **Measures of central tendency**
  Mean, Median, Mode
- **Measures of variability (dispersion)**
  Standard deviation, Variance, Range, Interquartile range

Measures of Variability

1. **Variance:**

$$\text{Sample variance} = s^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$$

2. **Standard deviation (SD):**

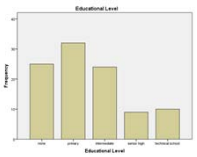
$$\text{Sample SD} = s = \sqrt{s^2}$$

3. **Range:**

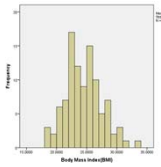
$$\text{Range} = \max - \min$$

Ways of Displaying Data

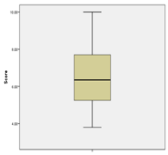
Bar chart



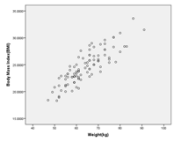
Histogram



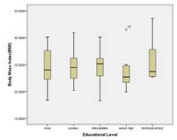
Box Plot



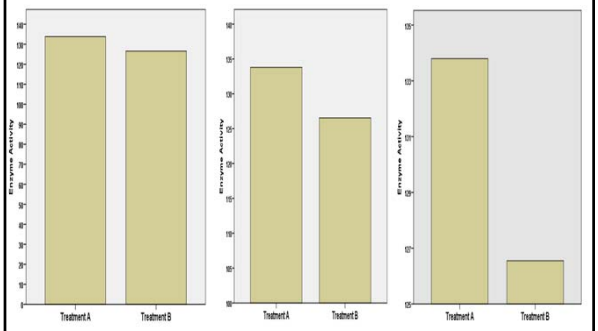
Scatterplot



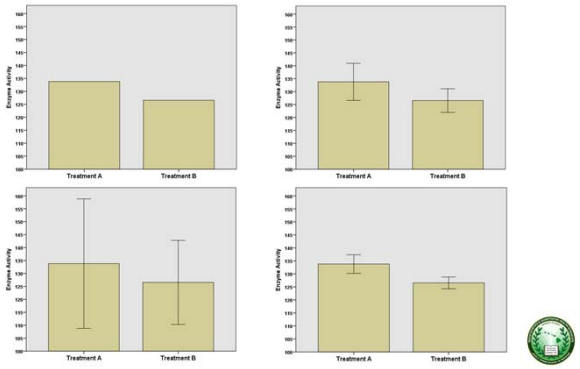
"Box and Bar" Plot



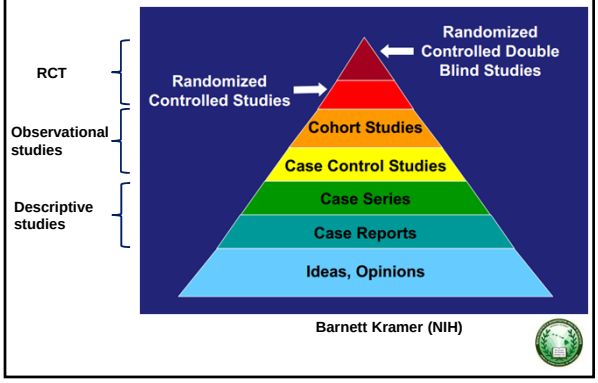
Different Scales



Different Error Bars



Clinical Research & Scientific Evidence



Barnett Kramer (NIH)

Tan et al. Long-term Survival Following Partial vs Radical Nephrectomy Among Older Patients With Early-Stage Kidney Cancer. *JAMA* 2012; 307:1629-1635.

Biomedical Research Process

Identifying a research question and a hypothesis

↓

Designing study and developing research protocol

↓

Gathering preliminary data and revising the protocol

↓

Conducting the study

↓

Analyzing data and interpreting results

Stats needed!

↓

Drawing conclusions and disseminating the results

Basic Principles of Experimental Design

- Replications
- Randomization
- Blocking (stratification)
- Blinding
- Factorial experiments

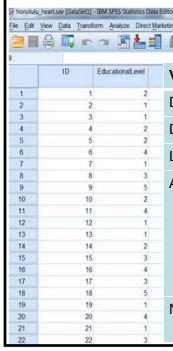
Handling A Confounding Variable (Z)

- If you can, fix a variable.
- If you can't, stratify it.
- If can't fix or stratify a variable, randomize it.

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \varepsilon$$

Data & Data Dictionary

SPSS: Honolulu Heart Study (partial data)



Variable	Education
Description/Label	Education Level
Data Type	Num – Categorical variable
Length	8
Allowable Values	1=none 2=primary 3=intermediate 4=senior high 5=technical school 6=university or above
Notes	Required field. No missing allowed.

Warning Signs

Data generation

Data consumption

MSCTR Curriculum

- BIOM 640 Introduction to Clinical Research (3 credits)
- BIOM 641 Legal & Regulatory Issues and Bioethics (2 credits; cross-listed with CMB626)
- BIOM 644 Translational Research Methods (2 credits)
- BIOM 645 Clinical Protocol Development (3 credits)
- BIOM 654 Medical Genetics (2 credits)
- QHS 601 Biomedical Statistics I (3 credits; cross-listed with TRMD 655)
- QHS 602 Biomedical Statistics II (3 credits)
- QHS 610 Bioinformatics I (3 credits; cross-listed with TRMD 653)
- QHS 611 Bioinformatics II (3 credits)
- QHS 620 Introduction to Clinical Trials (2 credits)
- QHS 621 Design and Analysis of Clinical Trials (2 credits)
- QHS 650 Secondary Data Analysis (2 credits)
- QHS 651 Secondary Data Analysis Practicum (2 credits)
- QHS 675 Biostatistical Consulting (2 credits)
- QHS 676 Biostatistical Consulting Practicum (1 - 2 credits)

MSCTR Graduate Program Website: msctr.jabsom.hawaii.edu

Collaboration with A Biostatistician

1. Early and often
2. Start the discussion when you have the initial idea
3. It is an iterative process
4. A collaborative effort: equal and fair
5. Ask questions so you can discuss about the general statistical approach without the statistician

JABSOM Office of Biostatistics & QHS

http://biostat.jabsom.hawaii.edu

NIH U54MD007601 **NIH U54MD007584**

Lecture Objectives

- Clarify the goal of statistics
- Grasp descriptive statistics
- Be familiar with various data presentation approaches
- Introduce key concepts of inferential statistics
- Survey some commonly used statistical approaches
- Understand basic research design principles
- Understand the pros and cons of large databases
- Build a foundation which will facilitate the active participation in clinical and translational research