HSCC 470
Research Methods and Data Analysis in Health Sciences

Descriptive Statistics I:

Central Tendency, Frequency Distribution, Sensitivity, Specificity, and Positive and Negative Predictive Value

Unit Objectives

Upon completion of this unit, the student will be able to:

- List the uses of descriptive statistics.
- Discuss the limitations of descriptive statistics.
- Describe a frequency distribution and histogram.
- Calculate the mean, median, and mode of a number array.
- Describe the uses of sensitivity, specificity, and predictive value in evaluating a clinical diagnostic test.

What are Statistics?

- **Statistics field involves methods for:**
  - Designing and carrying out research studies
  - Describing collected data
  - Making decisions, predictions, or inferences about phenomena represented by the data

What are Statistics? Continued

- Population vs. sample
- Parameters vs. statistics
- Descriptive vs. inferential

Frequency Distribution

- A listing of categories of possible values for a variable, together with a tabulation of the number of observations in each category.

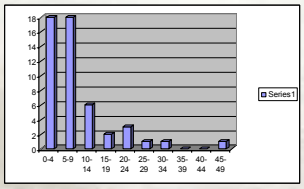
Raw Data

7	8	21
1	2	6
5	8	5
4	11	23
45	18	2
6	8	6
6	5	1
1	9	9
19	2	27
10	3	3
2	2	1
2	2	10
22	14	8
10	3	4
6	34	9
5	11	1
7	1	

Response Time	Frequency	Relative Frequency	Percentage
0-4	18	0.36	36.00%
5-9	18	0.36	36.00%
10-14	6	0.12	12.00%
15-19	2	0.04	4.00%
20-24	3	0.06	6.00%
25-29	1	0.02	2.00%
30-34	1	0.02	2.00%
35-39	0	0	0.00%
40-44	0	0	0.00%
45-49	1	0.02	2.00%

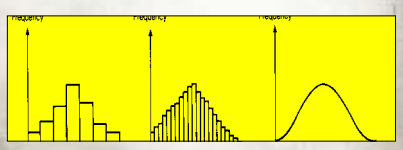
Frequency Distribution continued

- Histograms



Frequency Distribution continued

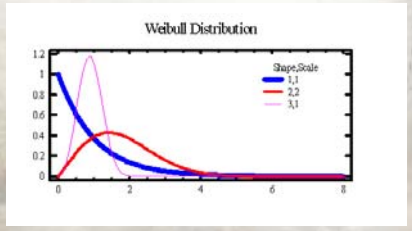
- **Sample and population distributions**



100 measurements 500 measurements Population

Frequency Distribution continued

- **Shapes of histograms**



Measures of Central Tendency

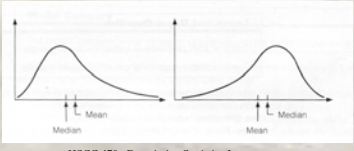
- **Notation**
  - n = sample size
  - Y_1, Y_2, Y_3….. Observation 1, observation 2…
  - $\bar{Y}$ (y bar) = sample mean
  - $\sum_{i=1}^{n} Y_i$ (sigma) = "sum of" values of Y between the first and the nth observation of Y

Measures of Central Tendency continued

- **Sample mean**
  - $\bar{Y} = \frac{Y1 + Y2 + Y3 \ldots Yn}{n}$
  - $\bar{Y} = \frac{\Sigma Y_i}{n}$

Measures of Central Tendency continued

- **Properties of the sample mean**
  - Only appropriate for data measured on at least an interval level
  - May be interpreted as the point of balance on the number line when an equal weight is placed at each measurement point.
  - Mean is "drawn" toward the tail of an asymmetrical distribution



Measures of Central Tendency continued

- **Sample Median**

  The measurement that falls in the middle when the sample data are ordered according to their magnitudes. If there is an odd number of observations, then the median is uniquely defined. If the sample size is even, there are two middle measurements and the median is the mean of the two.

Measures of Central Tendency Continued

- **Properties of the Median**
  - Median is appropriate for data measured on at least an ordinal scale.
  - For symmetrical distributions, the mean and median are the same.
  - For skewed distributions, the mean lies toward the direction of the skew (tail) relative to the median.

Measures of Central Tendency Continued

- **Percentile**

  The pth percentile is a number such that p% of the values of the distribution fall below it and (100-p%) lie above it.

Measures of Central Tendency continued

- **Sample mode**

  The value that occurs most frequently in the sample.
- **Properties of the Mode**
  - The mode is appropriate for all levels of measurement.
  - A frequency distribution is called bimodal, trimodal, etc., if there are two, three, etc., values that occur with the greatest frequency.
  - The mean, median, and mode are identical for a unimodal, symmetrical distribution, such as a bell-shaped distribution.

Sensitivity, Specificity, and Predictive Value

- Health care professional frequently use diagnostic tests to enhance their history-taking and physical exam in diagnosing disease and injury.
- There are no perfect diagnostic tests.
- Sensitivity, specificity, and predictive value are used to gauge the accuracy of diagnostic tests, triage criteria, physical signs and symptoms, etc.

Sensitivity, Specificity, and Predictive Value continued

		Disease: Present	Disease: Absent	
Test	Positive	*A*	*B*	*A + B*
	Negative	*C*	*D*	*C + D*
		A + C	*B + D*	*A+B+C+D*

A	correct
D	correct
B	false positive
C	false negative

Sensitivity = A/(A+C)
Specificity = D/(B+D)
Prevalence = (A+C)/(A+B+C+D)
Efficiency = (A+D)/(A+B+C+D)
Predictive Value = A/(A+B)

Sensitivity, Specificity, and Predictive Value continued

		Myocardial Infarction: Present	Myocardial Infarction: Absent	
Field Troponin	Positive	16	204	220
	Negative	6	252	258
		22	456	**478**

Sensitivity = A/(A+C)
Specificity = D/(B+D)
Prevalence = (A+C)/(A+B+C+D)
Efficiency = (A+D)/(A+B+C+D)
Predictive Value = A/(A+B)

Sensitivity = 73%
Specificity = 55%
Prevalence = 5%
Efficiency = 56%
Predictive Value = 73%

Clinical Application of Diagnostic Tests

- Trade-off of sensitivity vs. specificity
- Costs of tests for low prevalence conditions
- Diagnostic tests vs. clinical presentation
- Timing of positive test results