HSCC 470
Research Methods and Data Analysis in Health Sciences

Data Preparation and Analysis

Unit Objectives

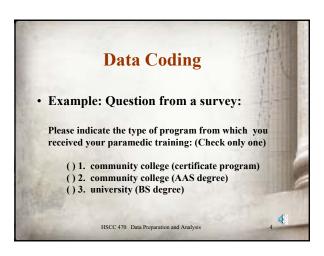**Upon completion of this unit, the student will be able to:**

- Describe the purpose of coding data
- Describe the process of coding data
- List the rules for coding data
- Correlate data codes and the level of measurement of data
- Describe how to deal with missing responses, incorrect responses, and "don't know" responses
- Discuss how to develop a code book
- List and describe the various data coding formats
- List and describe the various means for entering data
- List and describe how to verify your data once they have been entered

---

Data Preparation and Analysis

- **Data Coding**
  - The process of converting data into a form that can be analyzed by computer.

- **Some data are quantitative and do not need to be coded (ex: heart rate)**

- **Other data need to be converted into quantitative information**

Data Coding

- **Example: Question from a survey:**

Please indicate the type of program from which you received your paramedic training: (Check only one)

( ) 1. community college (certificate program)
( ) 2. community college (AAS degree)
( ) 3. university (BS degree)

Data Coding continued

- **Other data are not amenable to computer analysis**

  – Open-ended questions
  – Some types of observation data

Examples of Coding

- **Indicate your gender: (Circle one)**

    A. Male   B. Female

- **1 = male 2 = female**

- **Nominal scale**

Examples of Coding continued

- Indicate your age: _____   actual age

- Indicate your age:

1. <20	1 = under 20
2. 20-30	2 = between 20 and 30
3. 31-40	3 = between 31 and 40
4. >40	4 = greater than 40

- Ordinal scale

Rules for Coding

- intuitive
- mutually exclusive
- collectively exhaustive
- when in doubt, create another category
- match coding with highest level of data measurement
- selection of statistical tests is limited by data type
- construct codebook BEFORE printing final draft of instrument
- keep it simple

Miscellaneous Codes

- missing answer   9 or 99 or 999, etc.

- incorrect response 8 or 88 or 888, etc.

- don't know   7 or 77 or 777, etc.

Codebook Construction

- **Variable name**
- **Variable format (text, integer, fixed)**
- **Description**
- **Data codes**
- **Column numbers (fixed format only)**

Variable Name	Variable Format	Description	Column Numbers
Age	Integer	Patient's age in years	1-3
Gender	Integer	1 = male	4
		2 = female	
L_name	Character	Patient's last name	5-20
Weight	fixed	Patient's weight to the nearest tenth Kg	21-25

Data Formats

- **Fixed format**

  – variable 1   columns 1-3
  variable 2   columns 4-8
  variable 3   columns 9-10
  variable 4   column   11

  – Example: variable 1 = 8, variable 2 = 23, variable 3 = 14,
            variable 4 = 5

    00800023145

Data Formats continued

- **Free form**

    8 23   14   5

- **Delimited**

    8,23,14,5

Data Entry

- **Data Transfer Sheet**
- **Edge-coded Questionnaire**
- **Optical Scanning**
- **Direct Data Entry**
  - telephone entry
  - computerized survey/tests

Data Cleaning

- **"Spot checks"**
- **"Mini Analysis"**
  - Frequency verification
  - Mathematical verification
  - Outliers
- **Complete verification**