TECH 646 Analysis of Research in Industry and Technology

PART IV

Analysis and Presentation of Data:

Data Presentation and Description; Exploring, Displaying, and Examining Data; Hypothesis Testing; Measures of Association; Multivariate Analysis; Presenting Insights and Findings

Ch. 15 Data Preparation and Description

Lecture note based on the text book and supplemental materials:

Cooper, D.R., & Schindler, P.S., *Business Research Methods* (12 th edition), McGraw-Hill/Irwin

Paul I-Hai Lin, Professor of ECET
http://www.etcs.pfw.edu/~lin
A Core Course for M.S. Technology Program
Purdue University Fort Wayne Campus

1

Data Preparation and Description

Learning Objectives … Understand

- The importance of editing the collected raw data to detect errors and omissions.
- How coding is used to assign number and other symbols to answers and to categorize responses.
- The use of content analysis to interpret and summarize open questions.
- Problems with and solutions for “don’t know” responses and handling missing data.
- The options for data entry and manipulation.

2

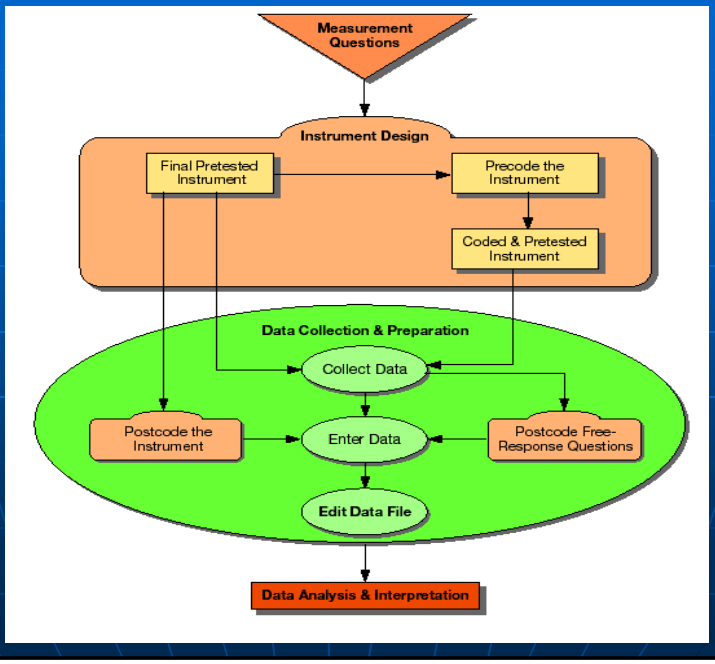
PulsePoint: Research Revelation

68

The percent of online consumers who put trust in online consumer recommendations.

3

Data Preparation in the Research Process



4

Research Thought Leader

"The goal is to transform data into information, and information into insight.

Carly Fiorina
former president and chairwoman, Hewlett-Packard Co

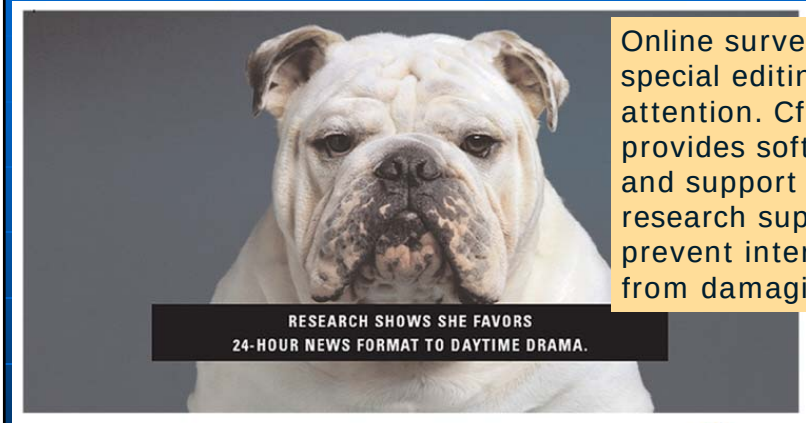
5

PulsePoint: Research Revelation

55 **The percent of white-collar workers who answer work-related calls or e-mail after work hours.**

6

Monitoring Online Survey Data



Now let's talk about software support. If the software hiccups or takes an unscheduled break, the wrong questions can get answered. Only nonstop support can keep that from happening.

That's why the world's top research organizations rely on CfMC Research Software, the only nonstop in the business.

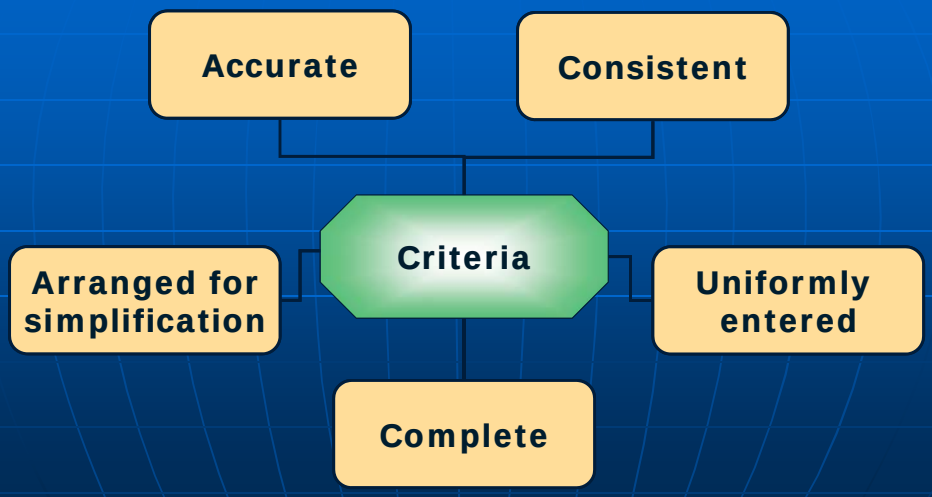
CfMC Research Software
Nonstop Support

www.cfmc.com San Francisco 866 LUV CfMC | New York 212.777.5120 | London 0-20 7837 5214

Online surveys need special editing attention. CfMC provides software and support to research suppliers to prevent interruptions from damaging data .

7

Editing (Purposes)



8

Types of Editing

- **Two Types**: Field editing and Central editing
- **Field Editing**
  - Field editing review (supervisor, soon after data collection)
  - Entry gaps identified
  - Callback made
  - Validate results (Re-interviewing at least 10% - typical amount)
- Central Editing

9

Field Editing

Field editing review
- **Entry gaps identified**
- **Callbacks made**
- **Results validated**

Speed without accuracy won't help the manager choose the right direction.

10

Central Editing

- Be familiar with instructions given to interviewers and coders
- Do not destroy the original entry
- Make all editing entries identifiable and in standardized form
- Initial all answers changed or supplied
- Place initials and date of editing on each instrument completed

11

CloseUp: Dirty Data (page 378)

- Problems with Dirty Data
  - Doctor's office: diagnose your illness or condition
  - Manufacturer: Try to identifying machine maintenance practices creating faulty parts
- Enterprise Information Systems
  - Data stores
  - Database
  - Data warehouse

12

CloseUp: Dirty Data (page 378)

- **Invalid: entry errors**
- **Incomplete: missing, siloed, turf wars**
- **Inconsistent: across databases**
- **Incorrect: lost, falsified, outdated**
- **Solutions: Data Steward, Data Protocols, Error Detection Software**

13

Snapshot: CBS Labs

- Las Vegas – research lab location
- CBS Television City Research Center in the MGM Grand Hotel and Casino; Nielsen TV viewership
- Advantages of : Touch screen data entry
- www.nielsen.com; www.viad.com

- **39 Million Visitors**
- **Show Screenings**
- **Dial Testing**
- **Surveys**
- **Focus Groups**

Coding

- **Coding** involves assigning "numbers" or "symbols" to answer so that the responses can be grouped into a limited number of categories.
- **Codebook**: coding scheme contains each variable in the study and specifies the application of coding rules to the variable.
- Codebook Construction
  - Spreadsheet, Minitab data file, SPSS data file

15

Exhibit 15-2 Sample Codebook of Questionnaires Items

Question	Variable Number	Code Description	Variable Name
_____	1	Record number	RECNUM
_____	2	Respondent number	RESID
1	3	5 digit zip code 99999 = Missing	ZIP
2	4	2 digit birth year 99 = Missing	BIRTH
3	5	Gender 1 = Male 2 = Female 9 = Missing	GENDER
4	6	Marital status 1 = Married 2 = Widow(er) 3 = Divorced 4 = Separated 5 = Never married 9 = Missing	MARITAL
5	7	Own–Rent 1 = Own 2 = Rent 3 = Provided 9 = Missing	HOUSING

Exhibit 15-3 Sample Questionnaires Items (Precoding)

1. What is the zip code of your residence? _ _ _ _ _

2. What is the year of your birth? 19_ _

3. Gender (1) Male
(2) Female
Indicate your choice → by number _

4. What is your marital status?
(1) Married
(2) Widow(er)
(3) Divorced
(4) Separated
(5) Never married
Indicate your choice → by number _

5. Do you own or rent your primary residence?
(1) Own
(2) Rent
(3) Living quarters provided
Indicate your choice → by number _

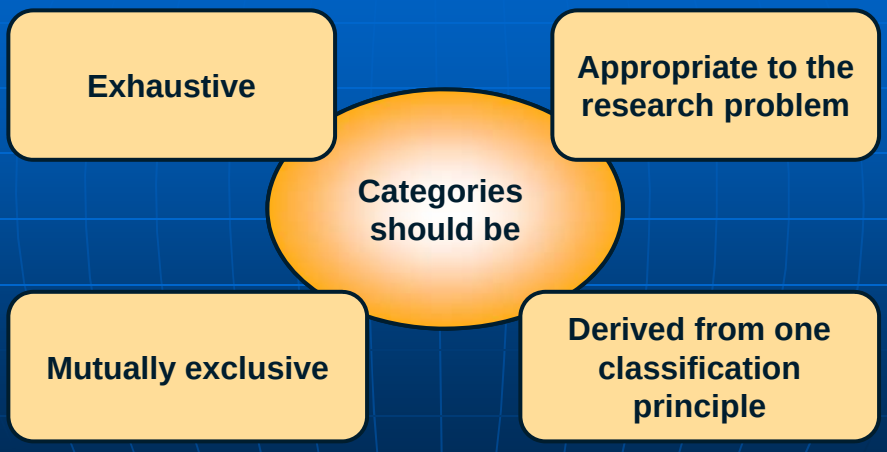
Exhibit 15-3 Coding Open-Ended Questions

6. What prompted you to purchase your most recent life insurance policy?

_____
_____
_____
_____
_____
_____
_____
_____

	Reason for purchase 1 = Mentioned 0 = Not mentioned	
8	Bought home	HOME
9	Birth of child	BIRTHCHD
10	Death of relative or friend	DEATH
11	Promoted	PROMO
12	Changed job/career	CHGJOB
13	Paid college expenses	COLLEXP
14	Acquired assets	ASSETS
15	Retired	RETIRED
16	Changed marital status	CHGMAR
17	Started business	STARTBUS
18	Expanded business	EXPBUS
19	Parent's influence	PARENT
20	Contacted by agent	AGENT
21	Other	OTHER

Coding Rules



19

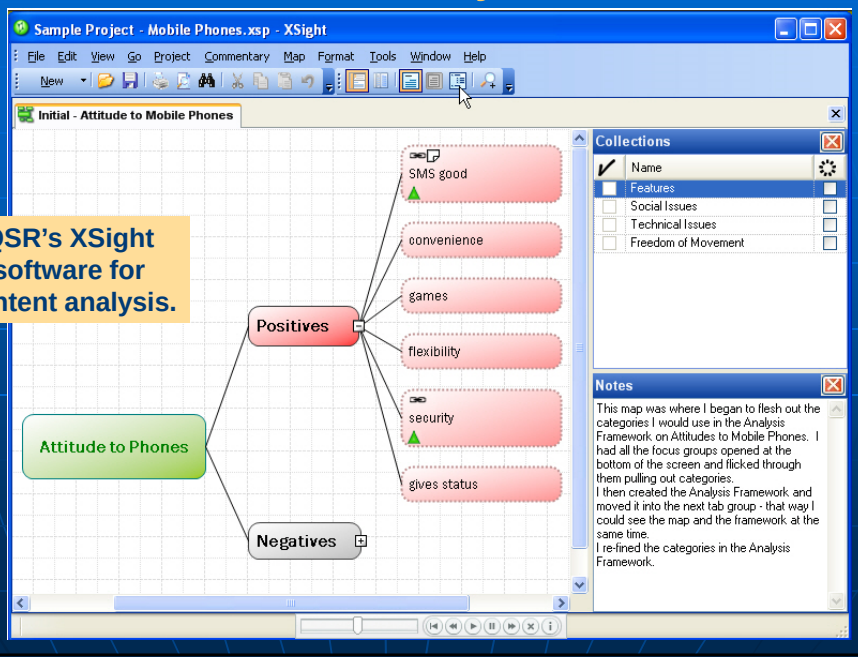
Content Analysis

- Measures the semantic content or what aspect of a message.
- It is used for open-ended questions
- Software for Content Analysis

20

Content Analysis

QSR's XSight software for content analysis.



Types of Content Analysis

- **Syntactical units**
  - Words, phrases, sentences, or paragraphs
- **Referential units** – described by
  - Words, phrases, and sentences, and
  - May be objects, events, persons, etc.
- **Propositional units**
  - Assertions about an object, event, or person.
- **Thematic units**
  - Topics contained within and across texts

22

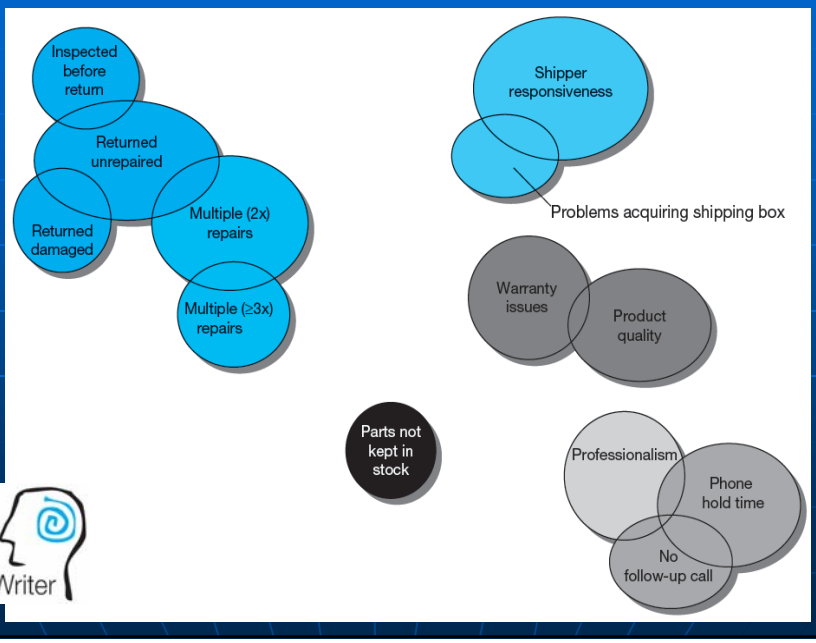
Exhibit 15-4 Open-Question Coding (before revision)

- **Question: "How can company-customer relations be improved?"**

Locus of Responsibility	Mentioned	Not Mentioned
A. Company	_____	_____
B. Customer	_____	_____
C. Joint Company-Customer	_____	_____
F. Other	_____	_____

Locus of Responsibility	Frequency (*n* = 100)
A. Management	
1. Sales manager	10
2. Sales process	20
3. Other	7
4. No action area identified	3
B. Management	
1. Training	15
C. Customer	
1. Buying processes	12
2. Other	8
3. No action area identified	5
D. Environmental conditions	
E. Technology	20

Proximity Plot
WindWriter Customer Complaints



Handling "Don't Know" Responses

Question: Do you have a productive relationship with your present salesperson?

Years of Purchasing	Yes	No	Don't Know
Less than 1 year	10%	40%	38%
1 – 3 years	30	30	32
4 years or more	60	30	30
Total	100% *n* = 650	100% *n* = 150	100% *n* = 200

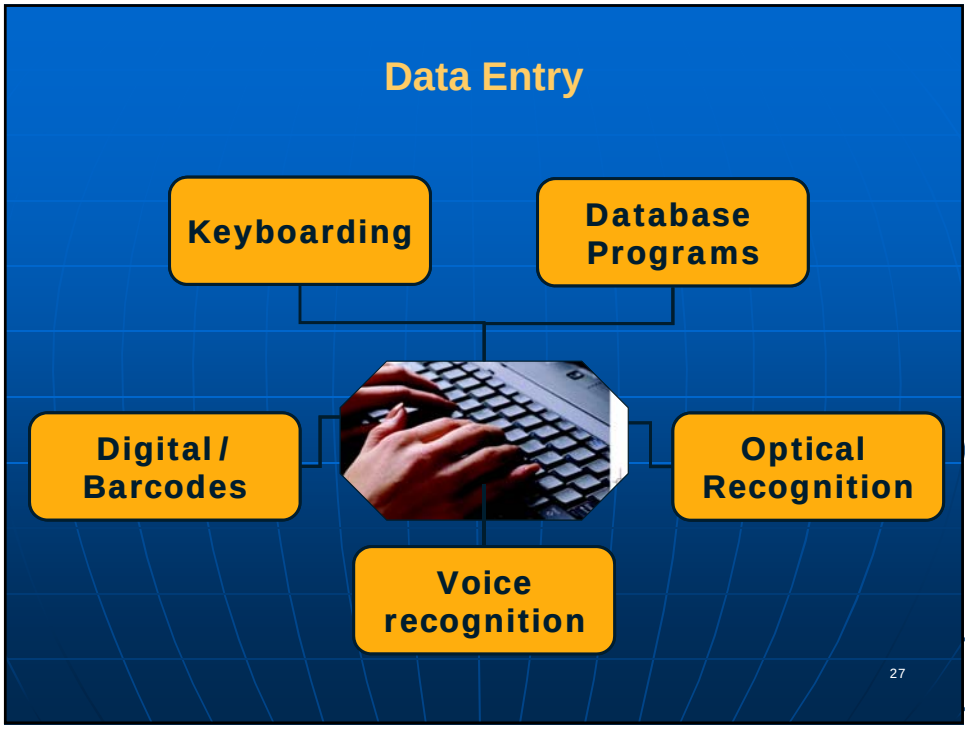
25

Don't Know Responses

- 1. Legitimate DK response when the respondent doe not know the answer
  - Meet research objectives (as expected and useful)
  - Awareness test
- 2. Researcher's failure to get the appropriate information

26

Data Entry



27

Missing Data

- Three basic types of missing data
  - Data missing completely at random (MCAR)
  - Data missing at random (MAR)
  - Data missing but not missing at random (NMAR)
- Three basic techniques for dealing with missing data
  - List-wise deletion
  - Pair-wise deletion
  - Replacement of missing values with estimated scores

28

Missing Data

Listwise Deletion

Pairwise Deletion

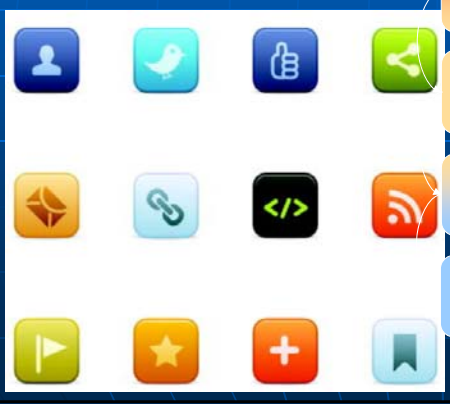
Replacement

Microsoft Excel - Sample Survey Spreadsheet

	C	D	E	F	G	H	I	J	K	L
	Q2	Q3	Q4	Q5	Q6a	Q6b	Q6c	Q6d	Q6e	Q6f
	2	1	10	2	1	2	1	1	4	4
	5	2	7	1	2	2	3	2	4	5
	2	1	6	2	2	4	3	4	4	4
	2	1	1	1	3	4	4	4	5	4
	8	2	8	2	3	5	4	2	5	1
	1	2	8	2	3	5	2	2	3	1
	3	1	8	1	2	5	3	5	3	3
	4	2	5	2	3	3	4	5	1	3
	2	1	9	1	3	2	4	5	2	5
	2	2	9	2	4	2	5	5	3	5
	5	2	9	1	4	1	1	3	1	5
	2	1	9	1	2	2	2	3	2	2
	1	2	3	2	5	3	3	4	2	1
	6	1	2	2	3	4	4	5	5	2
	4	2	3	1	1	4	3	1	5	3
	3	2	4	2	5	5	5	2	5	4
18	0017	1	3	1	6	1	5	5	2	1	1	4
19	0018	2	3	2	5	2	5	5	2	2	2	3

Snapshot: Seeking Clean Netnography Data

- Zdnet.com, Bizrate.com, Amazon.com, eBay.com, Elance.com, Complaints, reviewcentre.com, Epinion.com



Posted on Internet & intranets

Product & company reviews

Employee experiences

Message board posts

Discussion forum posts

30

Appendix 15a

Describing Data Statistically

31

Frequencies
(LCD TV Sales Statistics)

A

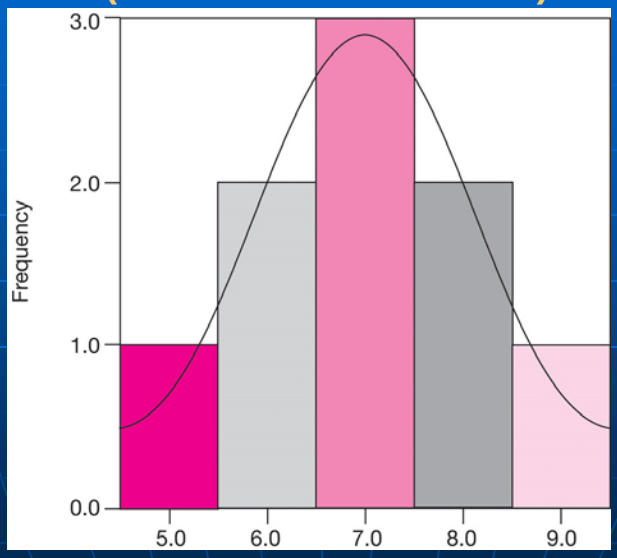
Unit Sales Increase (%)	Frequency	Percentage	Cumulative Percentage
5	1	11.1	11.1
6	2	22.2	33.3
7	3	33.3	66.7
8	2	22.2	88.9
9	1	11.1	100
Total	**9**	**100.0**	

B

	Unit Sales Increase (%)	Frequency	Percentage	Cumulative Percentage
Origin, foreign (1)	6 7 8	1 2 2	11.1 22.2 22.2	11.1 33.3 55.5
Origin, foreign (2)	5 6 7 9 **Total**	1 1 1 1 **9**	11.1 11.1 11.1 11.1 **100.0**	66.6 77.7 88.8 100.0

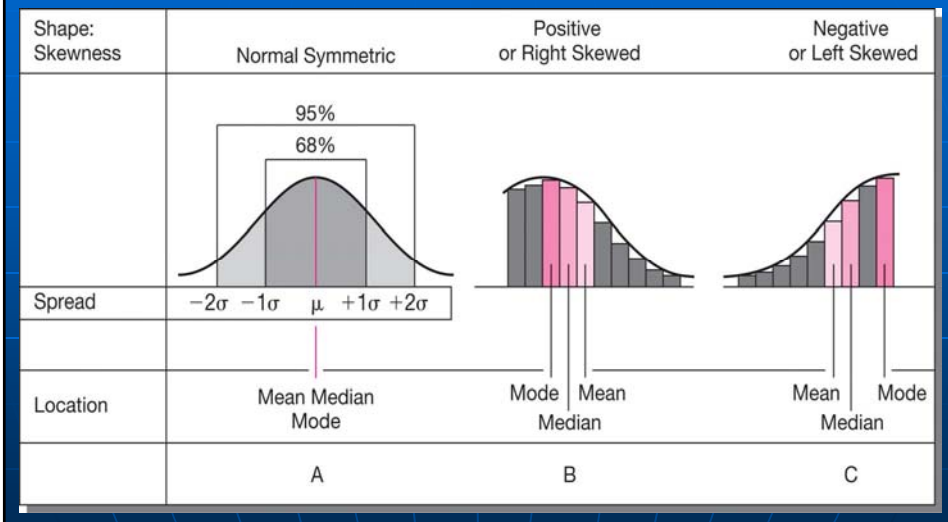
32

Distribution
(LCD TV Sales Statistics)



33

Characteristics of Distribution



34

Measure of Central Tendencies

Mean

Median

Mode

35

Measure of Variability

Variance

Quartile deviation

Standard deviation

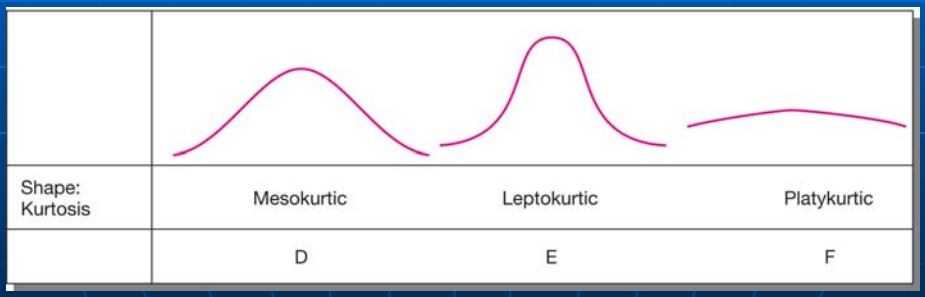
Interquartile range (1st/3rd difference, Midspread

Range

36

Summarizing Distribution Shape

- Kurtosis (ku)
  - Mesokutic: ku = 0; approach normal distribution
  - Leptokurtic: ku > 0
  - Platkutic: ku < 0



37

Symbols

Variable	Population	Sample
Mean	μ	X
Proportion	Π	p
Variance	σ^2	s^2
Standard deviation	σ	s
Size	N	n
Standard error of the mean	σ_X	S_X
Standard error of the proportion	σ_p	S_p

38

Summary