Study Guide
Marketing Research Test 4

Chapter 10

1. **Purpose of validating quantitative data**
    a. The purpose of validation is to determine if surveys, interviews, and observations were conducted correctly and free of bias or errors, and if data from other sources is accurate and consistent.
        i. **Curbstoning**: Cheating or falsification in the data collection process.
            1. Because of the potential for such falsification, data validation is an important step when the data acquisition process involves interviewers.

2. **Fraud in the data validation process**

    The process of validation covers 5 areas:

    a. Fraud - Was the person actually interviewed or was the interview falsified? Did the interviewer contact the respondent simply to get a name and address, and then proceeded to fabricate responses? Did the interviewer use a friend to obtain the necessary information?
        i. What information might we look for from respondents to detect fraud?
            1. For example: If you are stationed in Georgia, and one of your batches of surveys was answered by respondents from all over the U.S.A., then the person conducting the surveys probably faked the responses.
    b. Screening – To ensure accuracy of the data collected, respondents are screened according to some preselected criteria such as household income level, recent purchase of a specific product or brand, brand or service awareness, gender, or age. You need to make sure the people answering your survey are part of your population of interest and are a qualified respondent.
        i. Ex: Only female head of household with income of $25,000 or more.
    c. Procedure – In many marketing research projects, data must be collected according to a specific procedure.
        i. Ex: Customer exit interviews typically must occur in a designated place as the respondent leaves a retail establishment. You need to ensure that the interview took place at the proper setting, and not at a social gathering area or a park for example.
        ii. Certain procedures exist to make sure there is some data recall. If you wait to look to conduct an exit interview, customers may forget their experience.
    d. Completeness – Similar to fraud, but people may only answer the first few questions, then leave. For example, in order to speed through the data collection process an interviewer may ask the respondent only a few of the questions then make up answers to the remaining questions.
        i. Problem common to mall intercept

ii. To determine if the interview is valid, the researcher could recontact a sample of respondents and ask about questions from different parts of the questionnaire.
iii. This is not a problem with online surveys, because controls prevent respondent from skipping questions, but these controls can cause respondents to stop completing the survey before finishing, particularly if the questions are unclear, difficult, or uninteresting

e. Courtesy – respondents should be treated with courtesy and respect during the interviewing process. Situations can occur however, where the interviewer may inject a tone of negativity into the interviewing process.
   i. To ensure positive image, respondent callbacks are common to determine if the interviewer was courteous.

3. The coding process

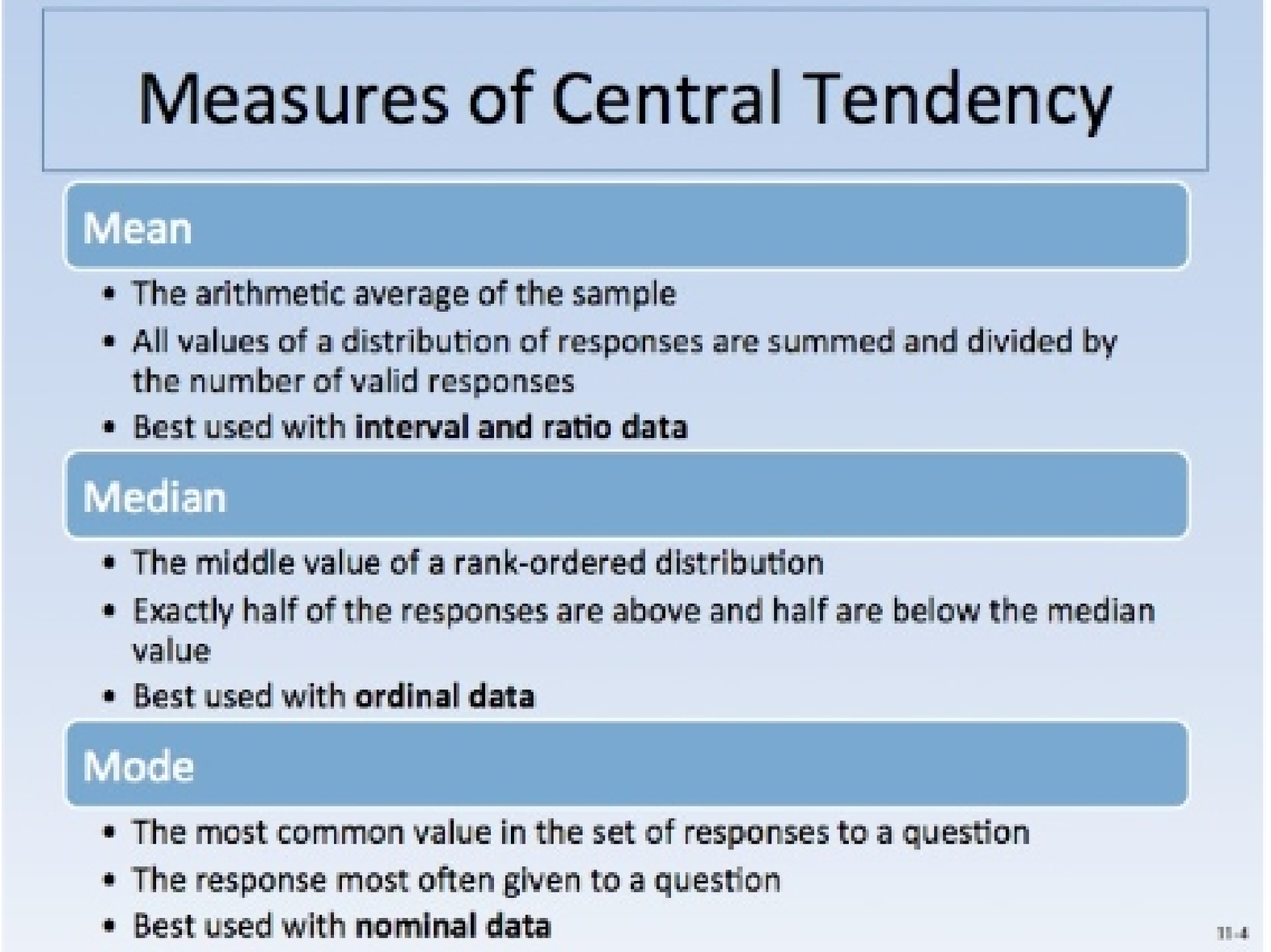
a. Grouping and assigning values to various responses from the survey instrument
   i. Codes are numerical
   ii. Can be tedious if certain issues are not addressed prior to collecting the data
b. Four-step process to develop codes for responses:
   i. Generate a list of as many potential responses as possible
   ii. Consolidate responses
   iii. Assign a numerical value as a code
   iv. Assign a coded value to each response

4. Types of data tabulation

a. Tabulation: is the simple process of counting the number of observations (cases) that are classified into certain categories
b. One-way tabulation: shows responses for a single variable. It is the categorization of single variables existing in a study.
   i. In most cases a one-way tabulation shows the number of responses (frequency count) who gave each possible answer to each question on the questionnaire.
   ii. They determine the amount of nonresponse to individual questions, and they can be used to locate mistakes in data entry
c. Cross-tabulation: Simultaneously compares two or more variables in the study; categorizing the number of respondents who have answered two or more questions consecutively, thus showing the relationship between those variables.
   i. Purpose is to determine if differences exist between subgroups of the total sample
   ii. A frequency distribution of responses on two or more sets of variables
   iii. Ex: it could show the number of males and females respondents who spent more than $7 at McDonalds versus who spent less.
   iv. Often used with nominal or ordinal scales
d. **One-way and cross-tabulations are descriptive statistics: used to summarize and describe the data obtained from a sample of respondents.**

Chapter 11

5. **Options for measuring central tendency**
    a. Every set of data collected needs some summary information developed that describes the numbers it contains
        i. Central tendency and dispersion
        ii. Relationships of the sample data
        iii. Hypothesis testing
    b. The mean, medium, median, and mode are measures of central tendency. These measures locate the central of the distribution. For this reason, the mean, median, and mode are sometimes called measures of location.



    c. Mean:
        i. Ex: Tells us the average number of cups of coffee the typical student drinks during finals to stay awake.
        ii. Shows some degree of central tendency, with most of the responses distributed close to the mean.
        iii. The mean is a very robust measure of C.T. (central tendency). It is fairly insensitive to data values being added or deleted.
            1. However, it can be subject to distortion if extreme values (outliers) are included in the distribution.
            2. **So the mean is best used with interval and ratio data, except when there are outliers. In this case, median and the mode are**