Chapter Fifteen

Data Processing And Fundamental Data Analysis

LEARNING OBJECTIVES

1. Get an overview of the data analysis procedure.
2. Understand validation and editing.
3. Learn how to code questions in surveys.
4. Understand the process of data entry so that information can be read by a computer.
5. Understand the importance of cleaning questionnaires so that they are free of errors.
6. Be familiar with tabulation and statistical analysis.
7. Gain insight into the graphic representations of data.
8. Comprehend descriptive statistics.

Overview of the Data Analysis Procedure

The Key Steps:

1	2	3	4	5
Validation & Editing	Coding	Data Entry	Logical Cleaning of Data	Tabulation & Statistical Analysis

Overview of the Data Analysis Procedure

Step One:

- ***Validation:*** Process of ascertaining that interviews actually were conducted as specified.
- ***Editing:*** Process of ascertaining that questionnaires were filled out properly and completely.
- ***Skip Pattern:*** Sequence in which later questions are asked, based on a respondent's answer to an earlier question or questions.

Step Two:

- ***Coding:*** Process of grouping and assigning numeric codes to the various responses to a question.

Step Three:

- ***Data Entry:*** Process of converting information to an electronic format.
- ***Intelligent Data Entry:*** Form of data entry in which the information being entered into the data entry device is checked for internal logic.
- ***Scanning:*** Form of data entry in which responses on questionnaires are read in automatically by the data entry device.

Skip Pattern

- “If MALE, go to question 3; if FEMALE, go to Question 4.”

Question

An interviewer, completed 50 questionnaires.

Further validation: Ten of the questionnaires.

A) Claimed age category was 30-40,

Marked category was 20-30.

B) "What is the most important problem facing our city government?"

1. **"The city council is too eager to raise taxes." (interviewer wrote)**
2. **In validation, "The city tax rate was too high." (response from the respondent)**

•As a valuator would you assume that these were honest mistakes and accept the entire lot of 50 interviews as valid? If not, what would you do?

Answer

- The error on the first questionnaire may have simply been a missed stroke of the pencil or computer.

- The second mistake, which is a completely different answer with similar content, gives cause for more concern.

- Check another ten questionnaires to see if more mistakes crop up. It may be that all the questionnaires should be validated.

Overview of the Data Analysis Procedure

Step Four:

- ***Clean the Data:*** Check for data entry errors or data entry inconsistencies.
- ***Logical or Machine Cleaning:*** Final computerized error check of data.
- ***Error Check Routines:*** Computer programs that accept instructions from the user to check for logical errors in the data.
- ***Marginal Report:*** Computer-generated table of the frequencies of the responses to each question, used to monitor entry of valid codes and correct use of skip patterns.

Step Five:

- ***Data Analysis –*** *e.g. One Way Frequency Tables:* Table showing the number of respondents choosing each answer to a survey question.
- ***Cross Tabulation Tables -*** Examination of the responses to one question relative to the responses to one or more other questions.

EXHIBIT 15.3 Sample of Responses to Open-Ended Question

Question: Why do you drink that brand of beer? (BRAND MENTIONED IN ANSWER TO PREVIOUS QUESTION)

Sample responses:

1. Because it tastes better.
2. It has the best taste.
3. I like the way it tastes.
4. I don't like the heavy taste of other beers.
5. It is the cheapest.
6. I buy whatever beer is on sale. It is on sale most of the time.
7. It doesn't upset my stomach the way other brands do.
8. Other brands give me headaches. This one doesn't.
9. It has always been my brand.
10. I have been drinking it for over 20 years.
11. It is the brand that most of the guys at work drink.
12. All my friends drink it.
13. It is the brand my wife buys at the grocery store.
14. It is my wife's/husband's favorite brand.
15. I have no idea.
16. Don't know.
17. No particular reason.

The Coding Process

EXHIBIT 15.4 Consolidated Response Categories and Codes for Open-Ended Responses from Beer Study

Response Category Descriptor	Response Items from Exhibit 14.1 Included	Assigned Numeric Code
Tastes better/like taste/tastes better than others	1, 2, 3, 4	1
Low/lower price	5, 6	2
Does not cause headache, stomach problems	7, 8	3
Long-term use, habit	9, 10	4
Friends drink it/influence of friends	11, 12	5
Wife/husband drinks/buys it	13, 14	6

QUESTION: How do you think the taste of Pepsi compares with Coca Cola?

Responses: The taste of Pepsi was

1) Better
2) Sweeter
3) Too sweet
4) Not as refreshing
5) More refreshing
6) Sweeter is better
7) Bolder taste
8) They taste about the same
9) I don't know
10) I don't have any idea
11) I don't drink soft drinks
12) Soft drinks are not good for you
13) Not a refreshing as Coca Cola
14) Coca Cola has always been my favorite brand

QUESTION: How do you think the taste of Pepsi compares with Coca Cola?

Responses: The taste of Pepsi was

1) Better
2) Sweeter
3) Too sweet
4) Not as refreshing
5) More refreshing
6) Sweeter is better
7) Bolder taste
8) They taste about the same
9) I don't know
10) I don't have any idea
11) I don't drink soft drinks
12) Soft drinks are not good for you
13) Not as refreshing as Coca Cola
14) Coca Cola has always been my favorite brand

One Way Frequency Tables

EXHIBIT 15.8 One-Way Frequency Table Using Three Different Bases for Calculating Percentages

Q.35 Why would you not consider going to Saint Paul for hospitalization?

	Total* Respondents	Total Asked	Total Answering
Total	300	64	56
	100%	100 %	100%
They aren't good/service poor	18	18	18
	6%	28 %	32%
Saint Paul doesn't have the services/equipment that Minneapolis does	17	17	17
	6%	27 %	30%
Saint Paul is too small	6	6	6
	2%	9 %	11%
Bad publicity	4	4	4
	1%	6 %	7%
Other	11	11	11
	4%	17 %	20%
Don't know/no response	8	8	
	3%	13 %	

*A total of 300 respondents were surveyed. Only 64 were asked this question because in the previous question those respondents said they would not consider going to Saint Paul for hospitalization. Only 56 respondents gave an answer other than "Don't know."

Cross Tabulations Examples

EXHIBIT 15.10 Sample Cross Tabulation

Q.30 If you or a member of your family were to require hospitalization in the future, and the procedure could be performed in Minneapolis or Saint Paul, where would you choose to go?

		Age			
	Total	18–34	35–54	55–64	65 or Over
Total	300	65	83	51	100
	100%	100%	100%	100%	100%
To a hospital in Saint Paul	144	21	40	25	57
	48.0	32.3%	48.2%	49.0%	57.0%
To a hospital in Minneapolis	146	43	40	23	40
	48.7%	66.2%	48.2%	45.1%	40.0%
Don't know/no response	10	1	3	3	3
	3.3%	1.5%	3.6%	5.9%	3.0%

Cross Tabulations Examples

EXHIBIT 15.11 Cross-tabulation Table with Column, Row, and Total Percentages*

Q.34 To which of the following towns and cities would you consider going for hospitalization?

	Total	Male	Female
Total	300	67	233
	100.0%	100.0%	100.0%
	100.0%	22.3%	77.7%
	100.0%	22.3%	77.7%
Saint Paul	265	63	202
	88.3%	94.0%	86.7%
	100.0%	23.6%	76.2%
	88.3%	21.0%	67.3%
Minneapolis	240	53	187
	80.0%	79.1%	80.3%
	100.0%	22.1%	77.9%
	80.0%	17.7%	62.3%
Bloomington	112	22	90
	37.3%	32.8%	38.6%
	100.0%	19.6%	80.4%
	37.3%	7.3%	30.0%

*Percentages listed are column, row, and total percentages, respectively.

Cross Tabulations Examples

EXHIBIT 15.11 Cross-tabulation Table with Column, Row, and Total Percentages*

Q.34 To which of the following towns and cities would you consider going for hospitalization?

	Total	Male	Female
Total	300	67	233
	100.0%	100.0%	100.0%
	100.0%	22.3%	77.7%
	100.0%	22.3%	77.7%
Saint Paul	265	63	202
	88.3%	94.0%	86.7%
	100.0%	23.6%	76.2%
	88.3%	21.0%	67.3%
Minneapolis	240	53	187
	80.0%	79.1%	80.3%
	100.0%	22.1%	77.9%
	80.0%	17.7%	62.3%
Bloomington	112	22	90
	37.3%	32.8%	38.6%
	100.0%	19.6%	80.4%
	37.3%	7.3%	30.0%

*Percentages listed are column, row, and total percentages, respectively.

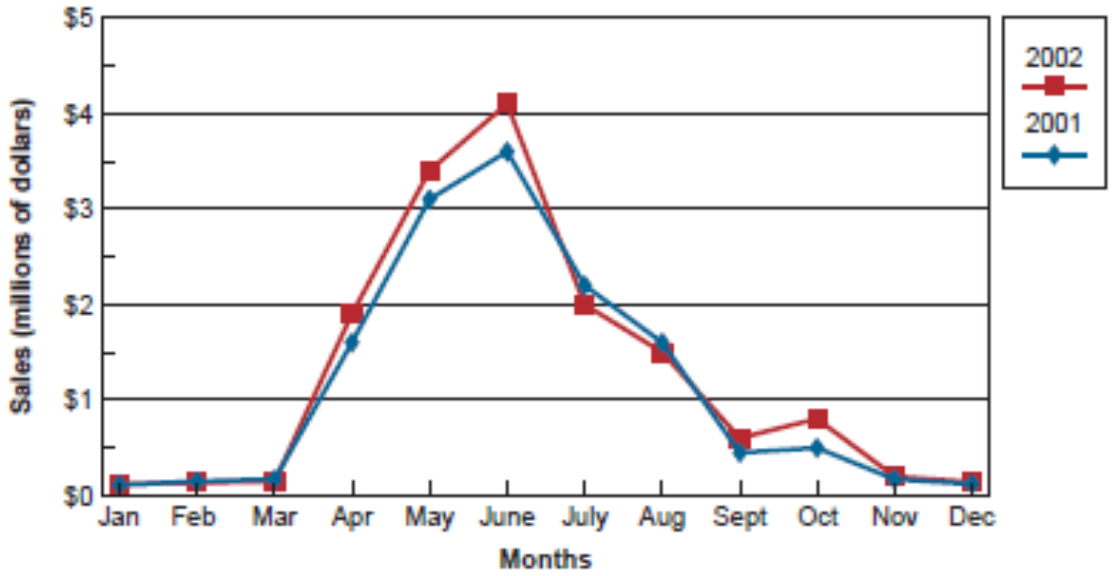
- Among all male respondents, 94%-Saint Paul, 79.1%-Minneapolis, 32.8% - Bloomington
- Among all female respondents, 86.7%-Saint Paul, 80.3%-Minneapolis, 38.6% - Bloomington

Graphic Representations of Data

Line Charts:

- Good for demonstrating linear relationships
- Particularly useful for presenting

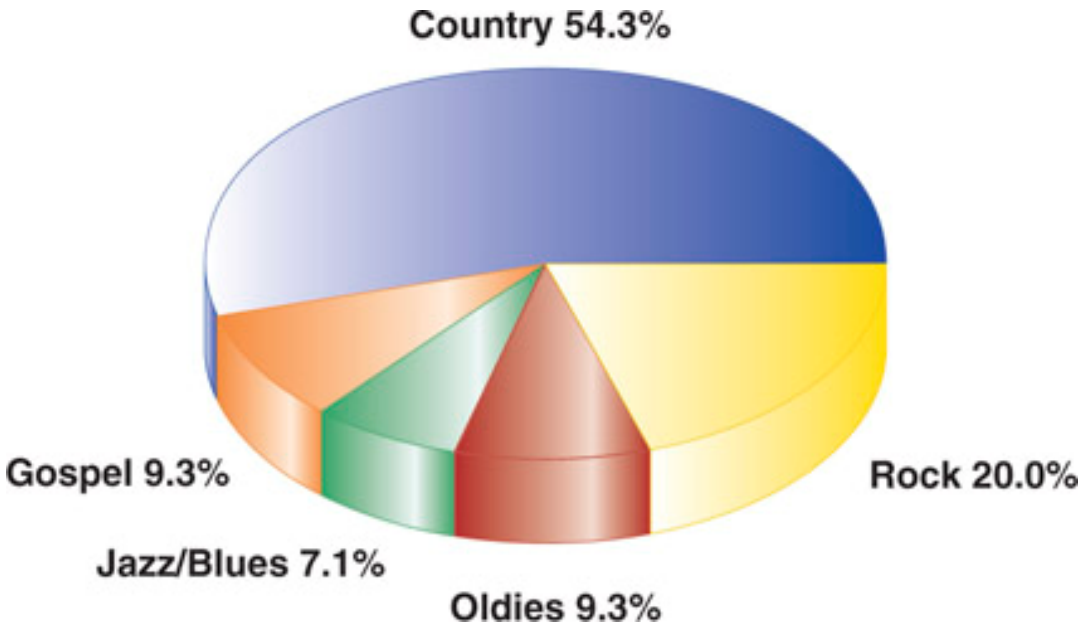
a given measurement taken at several points over time



Graphic Representations of Data

Pie Charts:

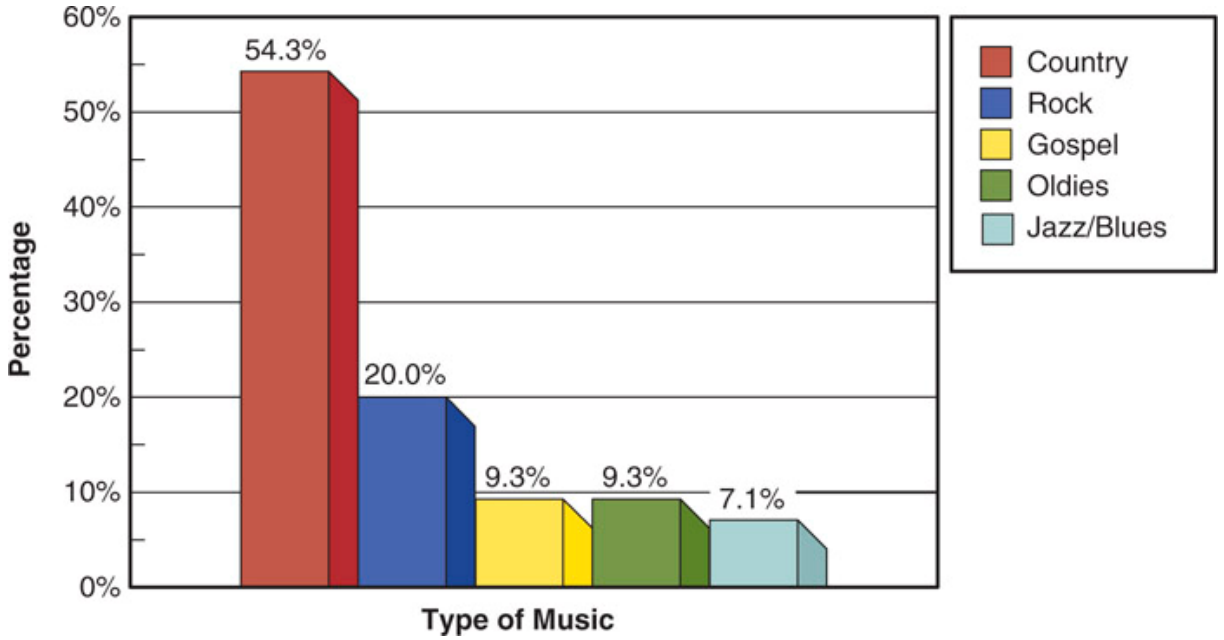
- Good for special relationships among data points
- Should total to 100%



Graphic Representations of Data

Bar Charts:

- Good for side by side relationships / comparisons
- Most flexible of the graphs



Descriptive Statistics

Measures of Central Tendency

Mean:

- The sum of the values for all observations of a variable divided by the number of observations.

Median:

- Value below which 50 percent of the observations fall.

Mode:

- The value that occurs most frequently.

Measures of Dispersion

These measures indicate how spread out the data are:

Variance:

- The sums of the squared deviations from the mean divided by the number of observations minus one.
- The same formula as standard deviation with the squaring.

Range:

- The maximum value for a variable minus the minimum value for that variable.

Standard Deviation:

- Measure of dispersion calculated by subtracting the mean of the series from each value in a series, squaring each result, summing the results, dividing the sum by the number of items minus 1, and taking the square root of this value.

Descriptive Statistics

Descriptive statistics are the most efficient means of summarizing the characteristics of large sets of data. In a statistical analysis, the analyst calculates one number or a few numbers that reveal something about the characteristics of large sets of data.

Significant discrepancies in "Mean" and Median" should cause you to look further into this data.

Years in Business	
Mean	22.4
Standard Error	2.6
Median	15.0
Mode	5.0
Standard Deviation	23.1
Sample Variance	534.5
Kurtosis	3.8
Skewness	2.1
Range	98.0
Minimum	2.0
Maximum	100.0
Sum	1770.5
Count	79.0