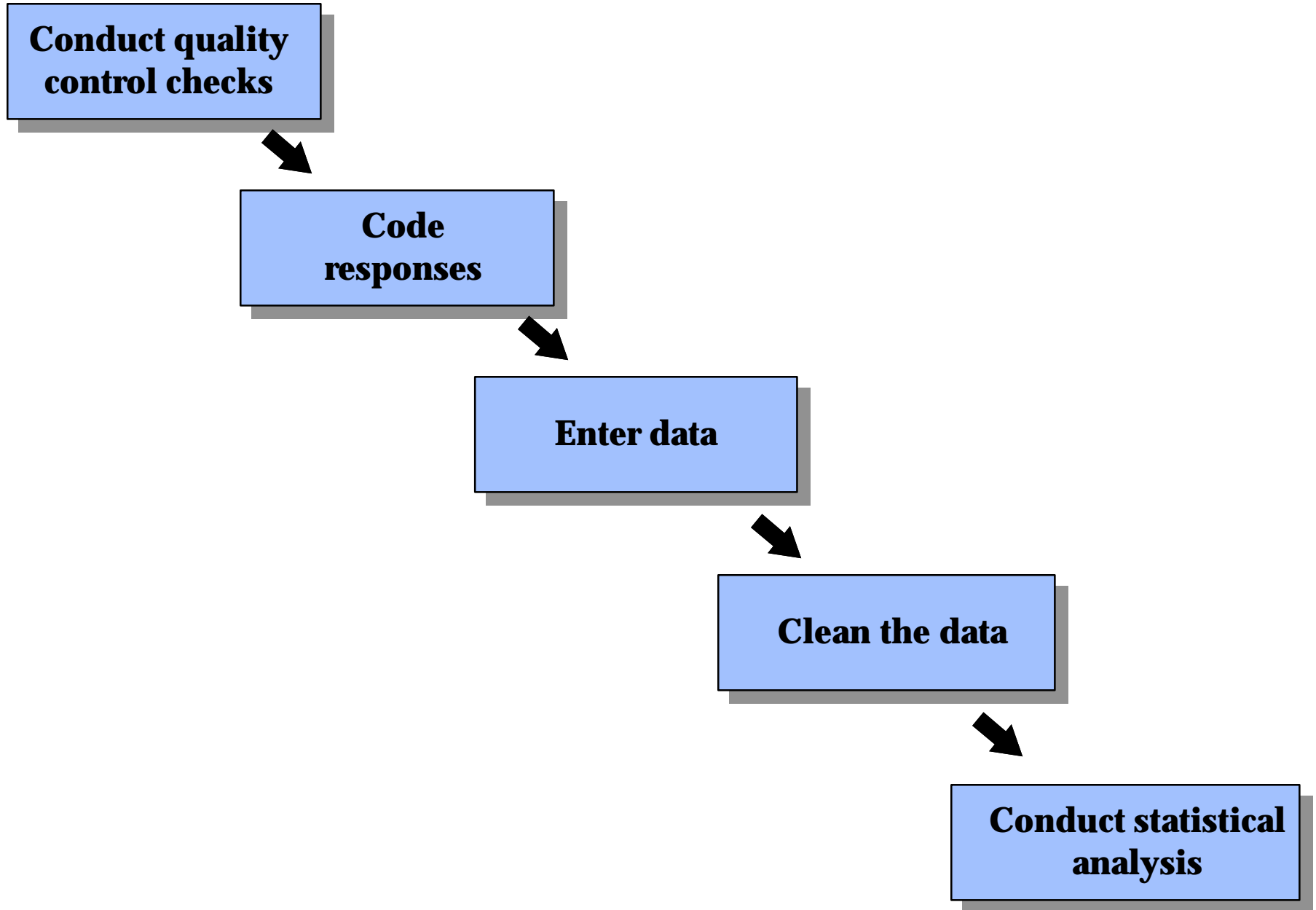


# The Data Analysis Process



## **The Coding Process for Open-Ended Questions**

**List responses**

```
graph TD; A[List responses] --> B[Consolidate responses  
(establish categories)]; B --> C[Set codes]; C --> D[Enter codes];
```

**Consolidate responses  
(establish categories)**

**Set codes**

**Enter codes**

**TABLE 15.1****Sample of Responses to Open-Ended Question**

---

*Question:* Why do you drink that brand of beer? (BRAND MENTIONED IN 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.
-

**TABLE 15.2****Consolidated Response Categories and Codes for Open-Ended Responses from Beer Study Introduced in Table 15.1**

---

<b>RESPONSE CATEGORY DESCRIPTOR</b>	<b>RESPONSE ITEMS FROM TABLE 15.1 INCLUDED</b>	<b>ASSIGNED NUMERIC CODE</b>
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
Don't know	15, 16, 17	7

---

## AN ILLUSTRATIVE CODEBOOK

Question number	Variable number	Card column	Format	Variable name	Category definitions
35	46	1/52	I1	Sex	1 = Female 2 = Male 3 = Missing data
36	47	1/53-54	I2	Age	Two-digit number 00-98 99 = Missing data
.	.	.	.	.	88 = Uncodable
.	.	.	.	.	.
74	121	3/21-23	F3.2	Grade point average	Three-digit number 000-400 with decimal place two places to left in field given

\* The use of "I" refers to an integer value in card column(s). The number after the "I" indicates the number of digits in the variable. For example, I1 indicates a one-digit integer variable, I6 a six-digit integer variable, and so on.

The use of "F" indicates that other variables in the card column(s) can take on real values. That is, the variable with an F can have a decimal place. The position of the decimal is given by the number after the decimal place, and the number of digits is given by the number before the decimal. For example, F3.2 indicates a real variable containing three digits with the decimal place position before the second digit. That is, the decimal is two places from the end of the field.

# A Typical Data Matrix

The concept of a data matrix.

		Variables (1 to $m$ )					
		Column numbers					
		1	2	.	.	.	$m$
<b>Cases (1 to <math>n</math>)</b>	1	2	625				4
	2	3	710				1
	3	1	521				3
	4	1	601				2
	5	4	706				1
	6	3	507				5
	7	2	429				1
	.						
	.						
	$n$	1	764				2

Variable 1 = Year in college

(1 = 1st year, 2 = 2nd year, 3 = 3rd year, 4 = 4th year)

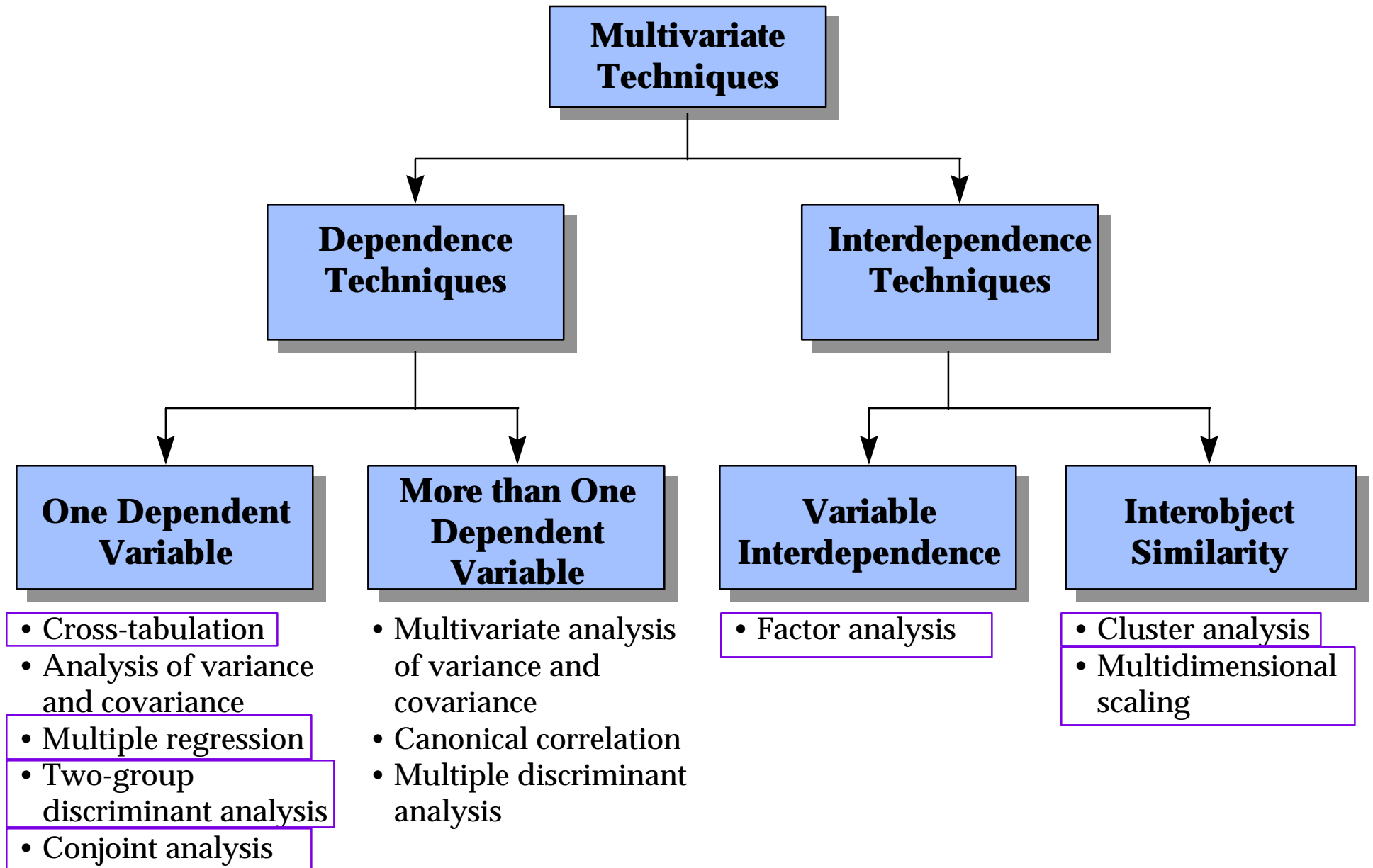
Variable 2 = SAT score (out of 800)

Variable  $m$  = Region of the country parents live in

(1 = Northeast, 2 = Southeast, 3 = Southwest, 4 = Northwest, 5 = Midwest)

# The Marketing Researcher's Bag of Tricks (Statistical Tool Kit)

Figure 16.7 A Classification of Multivariate Techniques



**TABLE 15.5**    **One-Way Frequency Table**

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 St. Paul, where would you choose to go?

	<b>TOTAL</b>
Total	300 100%
To a hospital in St. Paul	144 48.0%
To a hospital in Minneapolis	146 48.7%
Don't know/No response	10 3.3%

**TABLE 15.6**

**One-Way Frequency Table with Percents Shown for Total Respondents, Total Respondents Who Were Asked the Question, and Total Respondents Answering (gave response other than “don’t know”)**

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

	<b>TOTAL* RESPONDENTS</b>	<b>TOTAL ASKED</b>	<b>TOTAL ANSWERING</b>
Total	300 100%	64 100%	56 100%
They aren’t good/service poor	18 6%	18 28%	18 32%
St. Paul doesn’t have the services/equipment that Minneapolis does	17 6%	17 27%	17 30%
St. Paul is too small	6 2%	6 9%	6 11%
Bad publicity	4 1%	4 6%	4 7%
Other	11 4%	11 17%	11 20%
Don’t know/no response	8 3%	8 13%	

\* A total of 300 respondents were surveyed. Only 64 answered the question because in the previous question they said they would not consider going to St. Paul for hospitalization.

**TABLE 15.7****Percentages for a Multiple-Response Question Calculated on the Basis of Total Respondents and Total Responses**

---

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

---

	<b>TOTAL RESPONDENTS</b>	<b>TOTAL RESPONSES</b>
Total	300 100%	818 100%
Minneapolis	265 88.3%	265 32.4%
St. Paul	240 80.0%	240 29.3%
Bloomington	112 37.3%	112 13.7%
Rochester	92 30.7%	92 11.2%
Minnetonka	63 21.0%	63 7.7%
Eagan	46 15.3%	46 5.6%

---

## CROSS-TABULATIONS

**TABLE 17.5** Purchase of Fashion Clothing by Marital Status

Purchase of Fashion Clothing	CURRENT MARITAL STATUS	
	Married	Not Married
High	31%	52%
Low	69%	48%
Column total	100%	100%
Number of cases	700	300

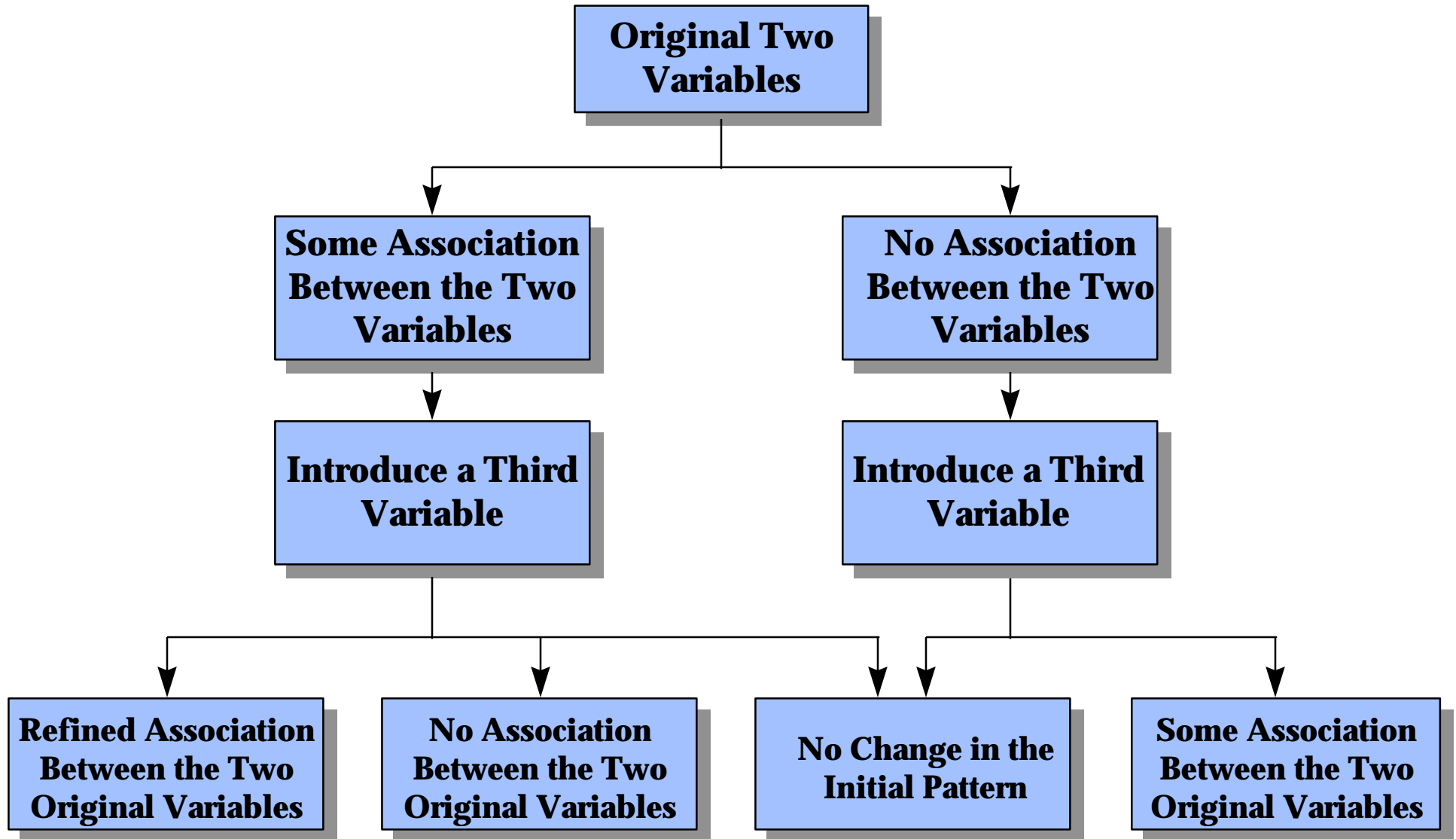
**Q. Can we say that marital status is related to the purchase of fashion clothing?**

**TABLE 17.6****Purchase of Fashion Clothing by Marital Status and Sex**

<b>Purchase of Fashion Clothing</b>	<b>SEX</b>			
	<b>MALE</b>		<b>FEMALE</b>	
	<b>MARITAL STATUS</b>		<b>MARITAL STATUS</b>	
	<b>Married</b>	<b>Not Married</b>	<b>Married</b>	<b>Not Married</b>
High	35%	40%	25%	60%
Low	65%	60%	75%	40%
Column total	100%	100%	100%	100%
Number of cases	400	120	300	180

**A. Yes, however effect appears to be stronger for females.**

**Figure 17.3 The Introduction of a Third Variable in Cross-Tabulation**



**TABLE 17.7****Ownership of Expensive Automobiles by Education Level**

<b>Own Expensive Automobile</b>	<b>EDUCATION</b>	
	<b>College Degree</b>	<b>No College Degree</b>
Yes	32%	21%
No	68%	79%
Column total	100%	100%
Number of cases	250	750

**Q. Can we say that education is related to the ownership of expensive automobiles?**

**TABLE 17.8****Ownership of Expensive Automobiles by Education and Income Levels**

<b>Own Expensive Automobile</b>	<b>INCOME</b>			
	<b>LOW INCOME</b>		<b>HIGH INCOME</b>	
	<b>EDUCATION</b>		<b>EDUCATION</b>	
	<b>College Degree</b>	<b>No College Degree</b>	<b>College Degree</b>	<b>No College Degree</b>
Yes	20%	20%	40%	40%
No	80%	80%	60%	60%
Column total	100%	100%	100%	100%
Number of cases	100	700	150	50

**A. No, it is income that appears to be related to the ownership of expensive automobiles.**

## Statistics Associated with Cross-Tabulation

### CROSS-TABULATION OF INCOME AND BRAND PURCHASED LAST

		<b>B</b> Brand purchased last			<b>Total</b>
		<b>B<sub>1</sub></b> Brand 1	<b>B<sub>2</sub></b> Brand 2	<b>B<sub>3</sub></b> Brand 3	
<b>A</b>	<b>Income</b>				
	A <sub>1</sub> Less than \$20,000	50	200	125	375
	A <sub>2</sub> \$20,000-\$40,000	200	100	350	650
	A <sub>3</sub> Over \$40,000	100	25	50	175
	Total	350	325	525	1200

**Is there a relationship between “income” and “brand purchased last”?**

#### SET UP HYPOTHESES

H<sub>0</sub>: (Called Null Hypothesis) → There is no relationship

H<sub>1</sub>: → There is a relationship

Chi-Square  $\rightarrow X^2 = \sum_{i=1}^R \sum_{j=1}^C \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$

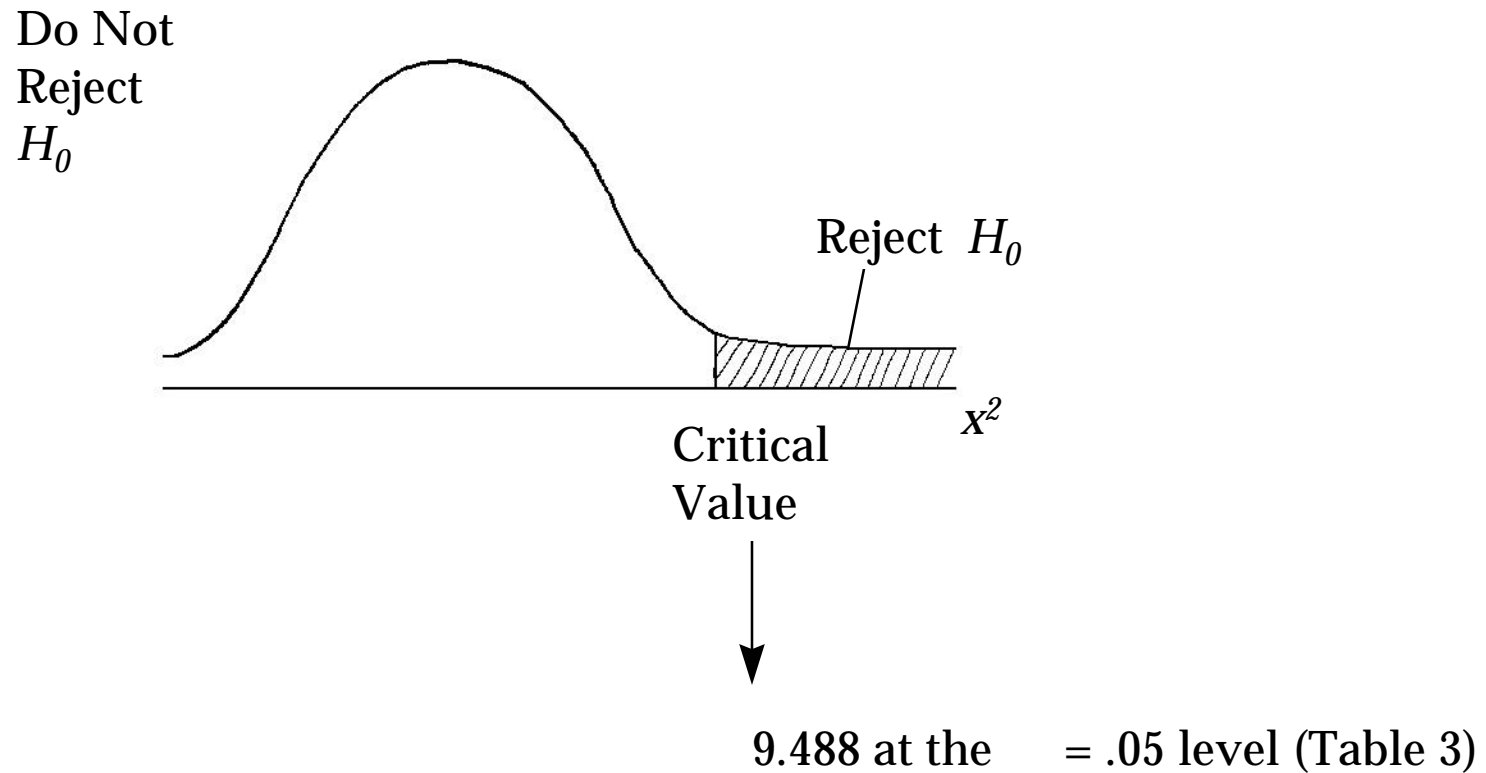
R  $\rightarrow$  Income Categories  
C  $\rightarrow$  Brand Last Purchased Categories

### CALCULATION OF BIVARIATE CHI SQUARE

Cell number	$O_{ij}$	$E_{ij}$	$O_{ij} - E_{ij}$	$(O_{ij} - E_{ij})^2$	$(O_{ij} - E_{ij})^2/E_{ij}$
1,1	50	109.38	-59.38	3526.0	32.2
1,2	200	101.56	98.44	9690.4	95.4
1,3	125	164.06	-39.06	1525.7	9.3
2,1	200	189.58	10.42	108.6	0.6
2,2	100	176.04	-76.04	5782.1	32.8
2,3	350	284.38	65.62	4306.0	15.1
3,1	100	51.04	48.96	2397.1	47.0
3,2	25	47.40	-22.40	501.8	10.6
3,3	50	76.56	-26.56	705.4	9.2
Total	1200	1200.00			252.2

$x^2 = 252.2$

## Figure 17.4 Chi-Square Test of Association

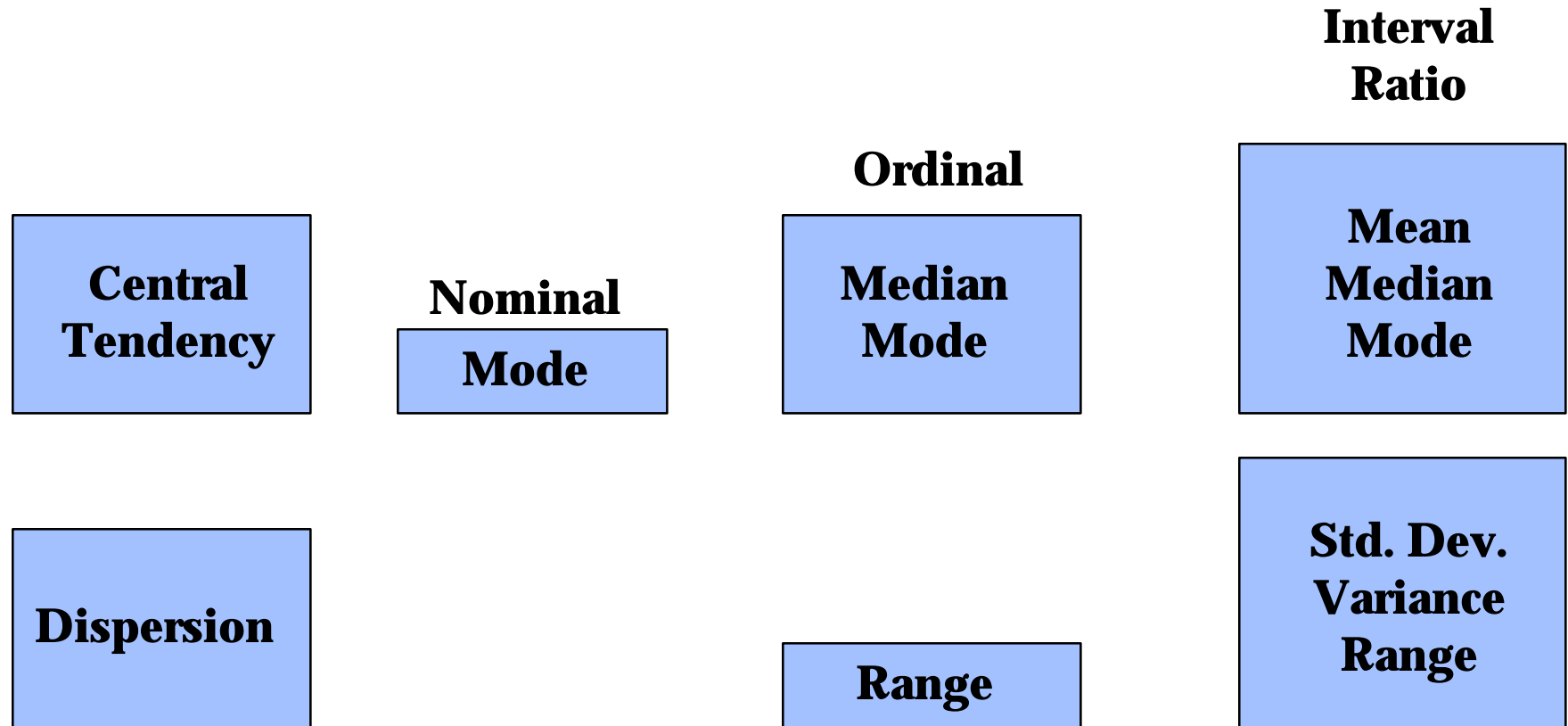


Calculated Value  $\longrightarrow$  252.2

Statistical Decision  $\longrightarrow$  Reject  $H_0$

MR conclusion ? (must always be in English)

# Legitimate Statistics for Different Levels of Measurement



**TABLE 15.11****Mean, Median, and Mode**

A total of 10 beer drinkers (drink one or more cans, bottles, or glasses of beer per day on the average) were interviewed in a mall intercept study. They were asked how many cans, bottles, or glasses of beer they drink in an average day. Their responses are summarized.

---

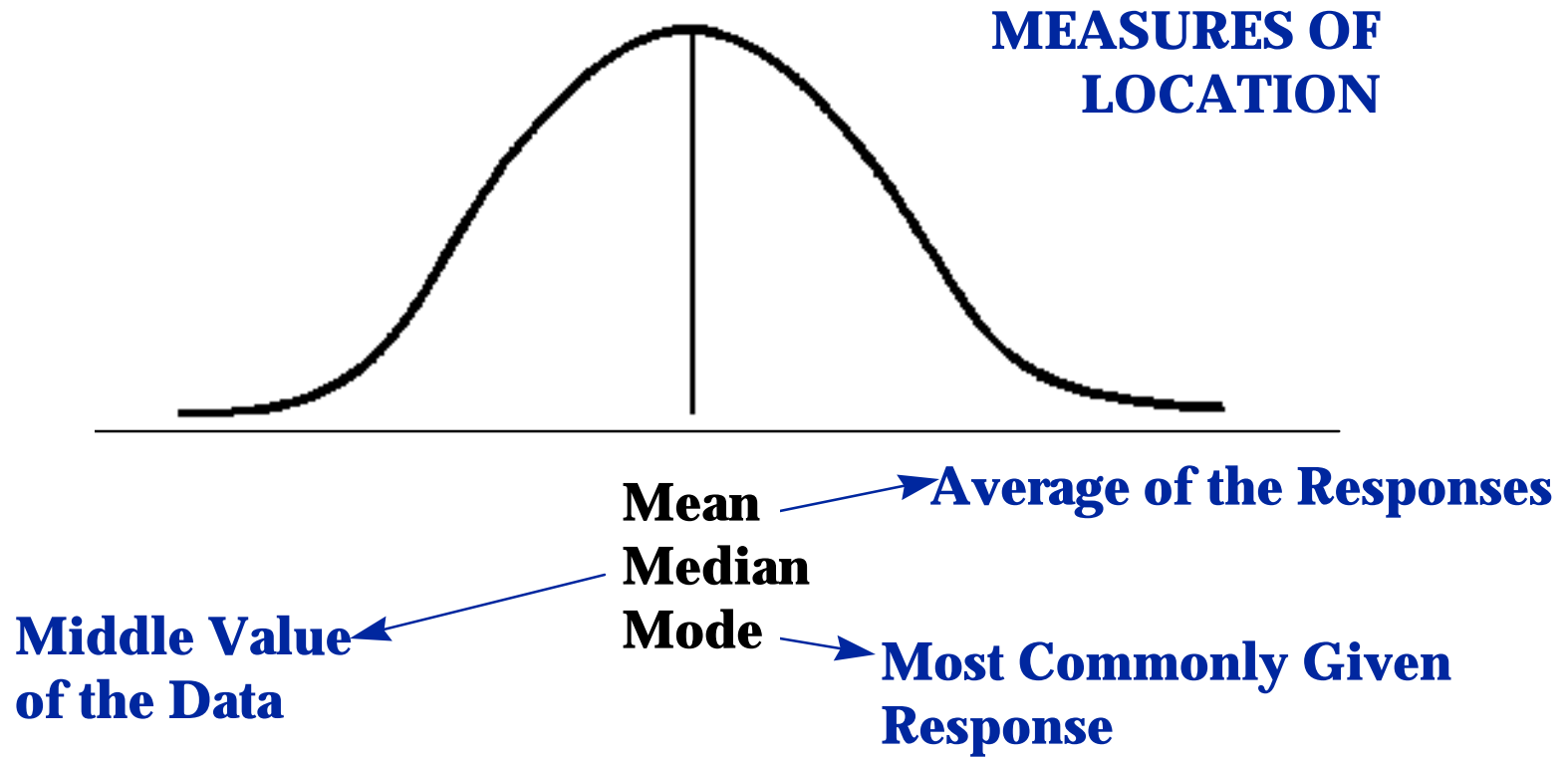
<b>RESPONDENT</b>	<b>NUMBER OF CANS/ BOTTLES/GLASSES PER DAY</b>
1	2
2	2
3	3
4	2
5	5
6	1
7	2
8	2
9	10
10	1

Mode = 2 cans/bottles/glasses  
Median = 2 cans/bottles/glasses  
Mean = 3 cans/bottles/glasses

---

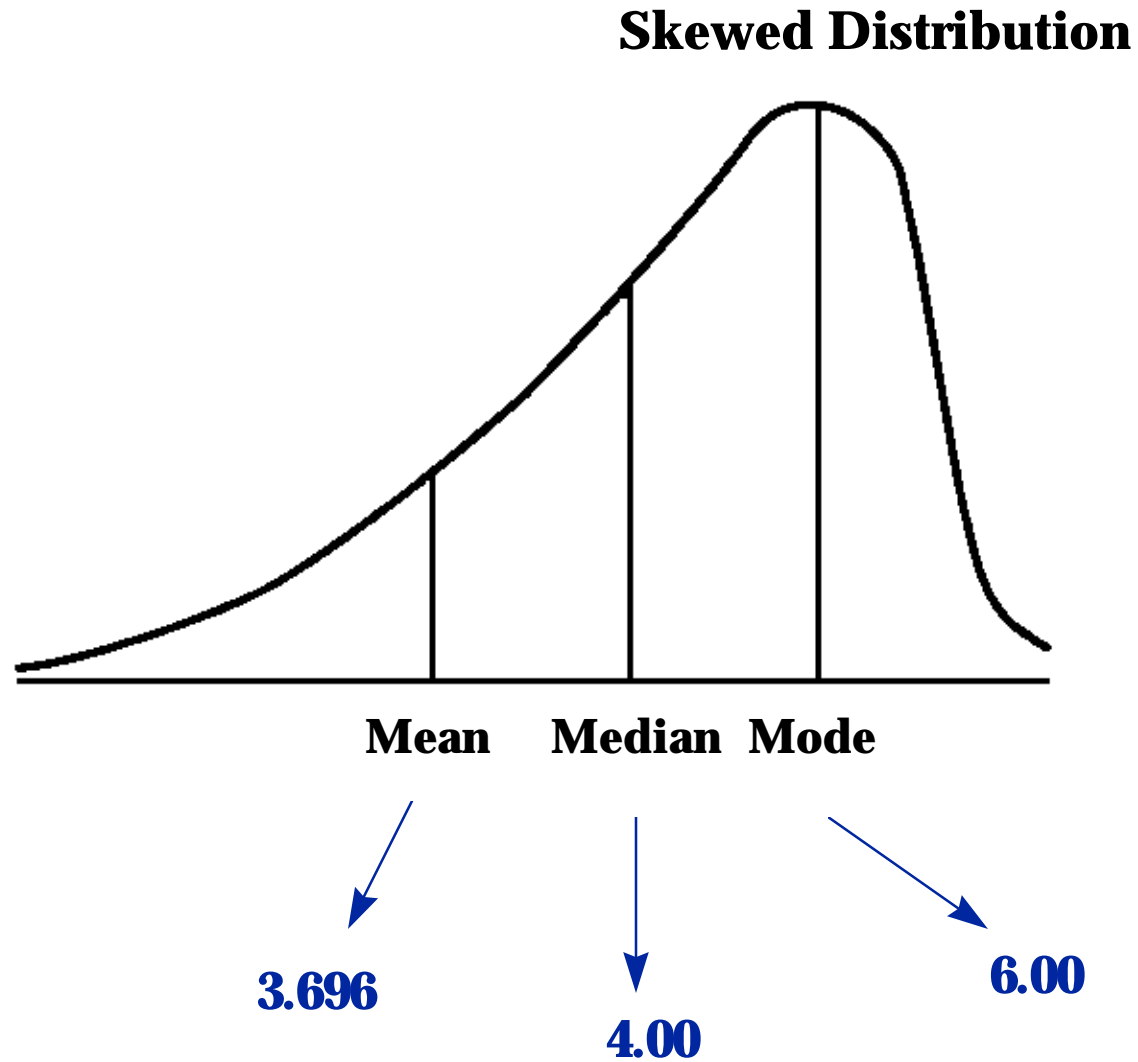
**Figure 17.2 Skewness of Distribution**

**Symmetric Distribution**



**(a)**

**Figure 17.2 Skewness of Distribution  
(continued)**



**(b)**

**TABLE 15.12****Measures of Dispersion and Measures of Central Tendency**

Consider a beer drinker example presented in Table 15.11. Assume that interviewing was conducted in two markets. The results for both markets are shown.

<b>RESPONDENT</b>	<b>NUMBER OF CANS/ BOTTLES/GLASSES MARKET ONE</b>	<b>NUMBER OF CANS/ BOTTLES/GLASSES MARKET TWO</b>
1	2	1
2	2	1
3	3	1
4	2	1
5	5	1
6	1	1
7	2	1
8	2	3
9	10	10
10	1	10
Mean	3	3
Standard Deviation	2.7	3.7

NOTE: Average beer consumption is the same in both markets---3 cans/bottles/glasses. However, the standard deviation is larger in Market Two, indicating more dispersion in the data. Whereas the mean suggests the two markets are the same, the added information provided by the standard deviation tells us they are different.

## Measures of Variability

**10**  
**15**  
**20**

**5**  
**10**  
**30**



**Spread is different, hence a measure of dispersion is necessary**

MEAN=15

MEAN=15

Std.Deviation in a sample of observations =  $S_x = \sqrt{\sum_{i=1}^n \frac{(X_i - \bar{X})^2}{n-1}}$

**STD. DEV = 5**

**STD. DEV = 13.2**

**COEFF. OF VARIATION = STD. DEV/MEAN**

1  
5  
10  
20  
70  
80  
120  
140  
150



**INTER  
QUARTILE  
RANGE = 110**

**RANGE = 149**