### **Treatment and Control Groups**

- The treatment group in an experiment is the group of sample members who receive the treatment being tested.
- The control group in an experiment is the group of sample members who do not receive the treatment being tested.

It is important for the treatment and control groups to be selected randomly and to be alike in all respects except for treatment.



### Definitions

- A case-control study is an observational study that resembles an experiment because the sample naturally divides into two (or more) groups.
- The participants who engage in the behavior under study form the cases.
- The participants who do not engage in the behavior are the controls.



## Basic Steps in a Statistical Study

- 1. State the goal of your study precisely.
- 2. Choose a representative sample from the population.
- 3. Collect data from the sample and summarize these data by finding sample statistics of interest.
- 4. Use the sample statistics to infer the population parameters.
- 5. Draw conclusions.

### **Elements of a Statistical Study**



Slide 5-4

### Definitions

The margin of error is used to describe a confidence interval that is likely to contain the true population parameter.

A confidence interval is from (sample statistic – margin of error) to (sample statistic + margin of error).

# Survey example

A national poll asked a random sample of 1081 adults whether they were satisfied with their jobs, and 54% said they were. The pollster concludes that 54% of adults in the U.S. are satisfied with their jobs, with a margin of error of 3% (at the 95% confidence level).

- The confidence interval is from 51% to 57%.
- A 95% confidence level means that the method used to find this margin of error results in a confidence interval that contains the actual population percentage about 95% of the time.

### Unit 5C

# Statistical Tables and Graphs

Copyright © 2011 Pearson Education, Inc.

Slide 5-7

### Data Types and Binning

Qualitative data describe qualities or categories.

Quantitative data represent counts or measurements.

When dealing with quantitative data categories, it is often useful to group, or **bin**, the data into class intervals that cover a range of possible values.



### Frequency Tables

A basic **frequency table** has two columns:

- The first column lists the categories of data.
- The second column lists the frequency of each category, which is the number of times each category appears in the data set.

Additional columns may include **relative frequency** (frequency expressed as a fraction or percentage of the total) or **cumulative frequency** (total of frequencies for the given category and all previous categories).

### Summarizing Raw Data

Consider the following 20 scores from a 100-point exam:

76 80 78 76 94 75 98 77 84 88 81 72 91 72 74 86 79 88 72 75

Determine appropriate bins and make a frequency table including columns for relative and cumulative frequency.

TABLE 5.3	Frequency Table for Binned Exam Scores						
Scores	Frequency	<b>Relative Frequency</b>	Cumulative Frequency				
95 to 99	1	0.05 = 5%	1				
90 to 94	2	0.10 = 10%	3				
85 to 89	3	0.15 = 15%	6				
80 to 84	3	0.15 = 15%	9				
75 to 79	7	0.35 = 35%	16				
70 to 74	4	0.20 = 20%	20				
Total	20	1.00 = 100%	20				

#### Slide 5-10

### **Bar and Pie Graphs**

A bar chart shows each category with a bar whose length corresponds to its frequency or relative frequency.

Pie charts are used primarily for relative frequencies, because the total pie must always represent the total relative frequency of 100%. The size of each wedge is proportional to the relative frequency of the category it represents.

### Bar and Pie Graphs



### Definitions

- A histogram is a bar graph for quantitative data categories. The bars have a natural order and the bar widths have specific meaning.
- A line chart shows the data value for each category as a dot, and the dots are connected with lines. For each dot, the horizontal position is the *center* of the bin it represents and the vertical position is the data value for the bin.

### Histogram and Line Chart

The histogram and line chart below both show the same data.



Copyright © 2011 Pearson Education, Inc.

### Example

minimum

### TABLE 5.1 Percent of Adult Population of Hispanic Origin, by State (2000)

State	Percent	State	Percent	State	Percent
Alabama	1.5	Louisiana	2.4	Ohio	1.9
Alaska	4.1	Maine	(0.7)	Oklahoma	5.2
Arizona	25.3	Maryland	4.3	Oregon	8.0
Arkansas	2.8	Massachusetts	6.8	Pennsylvania	3.2
California	32.4	Michigan	3.3	Rhode Island	8.7
Colorado	17.1	Minnesota	2.9	South Carolina	2.4
Connecticut	9.4	Mississippi	1.3	South Dakota	1.4
Delaware	4.8	Missouri	2.1	Tennessee	2.0
Florida	16.8	Montana	2.0	Texas	32.0
Georgia	5.3	Nebraska	5.5	Utah	9.0
Hawaii	7.2	Nevada	19.7	Vermont	0.9
Idaho	7.9	New Hampshire	1.7	Virginia	4.7
Illinois	10.7	New Jersey	13.3	Washington	7.2
Indiana	3.5	New Mexico	42.1	West Virginia	0.7
Iowa	2.8	New York	15.1	Wisconsin	3.6
Kansas	7.0	North Carolina	4.7	Wyoming	6.4
Kentucky	1.5	North Dakota	1.2	by volues are promoted	transfer if near

maximum

# Making a Histogram

### **Step 1.** Divide the range of the data into classes of equal width.

The data ranges from 0.7 to 42.1 percent. If we make the classes of width 5 percent, we will have nine classes between 0 and 45 percent. Since the data is specified to the tenth of a percent, we specify the classes to the same accuracy.

Class	Count	Class	Count	Class	Count
0.0 to 4.9	27	15.0 to 19.9	4	30.0 to 34.9	2
5.0 to 9.9	13	20.0 to 24.9	0	35.0 to 39.9	0
10.0 to 14.9	2	25.0 to 29.9	1	40.0 to 44.9	1

**Step 2.** Count the number of individuals in each class.

### Step 3. Draw the histogram



Horizontal scale displays the variable. The scale runs from 0 to 45, the span of the classes.